Network Theory Approaches for the Analysis of Ordered Data

Ryan Flanagan

Submitted in partial fulfilment of the requirements of the degree of
Doctor of Philosophy

School of Mathematical Sciences
Queen Mary University of London
United Kingdom

May 2020

Statement of Originality

I, Ryan Flanagan, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

Details of collaboration and publications:

- Chapter 2 details joint work with Lucas Lacasa.: L. Lacasa and R. Flanagan, "Time reversibility from visibility graphs of nonstationary processes". *Physical Review E*, 92(2):022817, 2015
- Chapter 3 details joint work with Lucas Lacasa: R. Flanagan and L. Lacasa. "Irreversibility of financial time series: A graph-theoretical approach". *Physics*

Letters A, 380(20):1689-1697, 2016.

- Chapter 4 detail joint work with Lucas Lacasa, Uri Hasson, Jacopo Iacovacci, et al: U. Hasson, J. Iacovacci, B. Davis, R. Flanagan, E. Tagliazucchi, H. Laufs, and L. Lacasa. "A combinatorial framework to quantify peak/pit asymmetries in complex dynamics". *Scientific Reports*, 8(1), 2018.

- Chapter 5 details joint work with Lucas Lacasa and Vincenzo Nicosia: R. Flanagan, L. Lacasa, and V. Nicosia. "On the spectral properties of Feigenbaum graphs". *Journal of Physics A: Mathematical and Theoretical*, 53(2):025702, 2020.

- Chapter 6 details joint work with Lucas Lacasa, Emma K. Towlson, Sang Hoon Lee et al: R. Flanagan, L. Lacasa, E. K. Towlson, S. H. Lee, and M. A. Porter. "Effect of antipsychotics on community structure in functional brain networks". *Journal of Complex Networks*, 7(6):932960, 2019.

Acknowledgments

I would like to thank my supervisor Lucas Lacasa for his extraordinary support, without which I would have not made it this far. I am also thankful for the support from the School of Mathematical Sciences (in particular Elisa Piccaro), the Research Degrees Office, and the EPSRC (for which this research was generously supported through DTP EP/N013492/1). I also do not forget the help from my undergraduate tutor, Frances Kirwan, and project supervisor, Mason Porter, for helping me start this journey.

I am fortunate to have a large number of friends, and a supportive family, whom I relied on during a rather turbulent time. In particular, my mother and my step father, and my father and my step mother, my two sisters, Lisa and Katy, my aunties and uncles, my friends Sam, Alex, Sumil, Lucy, Laura, Lily, Sammy, David, Filip, Abdulkadir, Scott, Neil, Satya and Carmel all provided practical support during that time. I also thank Jack Bartley for some template help. Lastly, I would like to thank Megan and Ørn for giving me a final push.

Abstract

Graphs are mathematical structures comprised of a set of nodes connected by edges, and network science is the application of graph theory to real world data. Networks are used as a model to analyse how entities, either individual actors, or complex systems, interact with one another.

The research here will consist of extracting networks (we will use the terms "graph" and "network" interchangeably) from ordered series, which we will focus on series ordered by time. We will either do this with the aid of the visibility graph, which is a method, based on visibility, for mapping a time series in to a graph, or through estimating the wavelet correlation, a more conventional method used in neuroscience. The aim is to describe the structure of time series and their underlying dynamical properties in graph-theoretical terms, and then using this motivation to analyse large data sets spanning several disciplines. We will describe a method, using the visibility graph, for quantifying reversibility of non-stationary processes and apply this method to a large financial data set, with the intent of ranking companies based on their irreversibility. We also use the visibility graph to develop a method which efficiently quantifies the asymmetries between minima and maxima in time series, and we then apply the method to a variety of data sets. Continuing with the theme of visibility, we study the spectral properties of visibility graphs extracted from trajectories of the logistic map undergoing a period-doubling route to chaos (known as the Feigenbaum scenario). Finally, we will use wavelet correlation to construct networks from fMRI time series, and examine community structure with the aim of differentiating between brain networks of patients with schizophrenia from control subjects. The general format throughout this thesis will start with theory, followed by extensive numerical simulations, which we can then apply the methods to real data sets.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The goal of this thesis is to investigate the structure of ordered series (for example, trajectories of a dynamical system, or real world fMRI data) by developing and using new tools which originate in network and graph theory (in what follows, we will use the words graph and network interchangeably). Here we will focus on series ordered by time, but in practice all of the methods and results are relevant to any series with a suitable ordering.

In recent years we have witnessed how both the theoretical and the experimental investigation of time series benefit from insights which originate in different disciplines such as graph theory and network science. Let us consider one scenario. In neuroscience it is customary to track and record brain activity by measuring the electrical activity in different regions of the brain. By installing sensors in "regions of interest" (RoIs), we can extract multivariate time series (where each entry corresponds to the electrical activity in a different RoI) which characterise the brain's electrical activity of a subject. We can do this while this subject is performing a given task, where the default, control task is the subject at rest (which is also called "resting state"). A similar approach is used with functional magnetic resonance imaging (fMRI) data. To investigate how different regions of the brain dynamically coordinate, and to have a good picture of

this coordination at a global scale, it is customary to construct "functional networks". These are graphs where each RoI corresponds to a different node in the network, and two nodes are linked by a weighted edge, where the weight corresponds to the degree of correlation or synchronisation between the two nodes (one can then turn this weighted network in to a binary network by thresholding). Once a functional network is extracted, then one can use such a network as a global descriptor of the multivariate signal. For instance, one can investigate whether the effect of a drug administered to the subject is inherited in the structure of the functional networks in a quantitative way. Similarly, one can investigate whether different psychological disorders, such as schizophrenia, can be "characterised" in such functional networks, and if so, whether these can be used for diagnoses. In Chapter 6 we will investigate how such an approach can help us to understand the effect of different antipsychotic drugs in the treatment of schizophrenia. To do that, we will first construct functional networks extracted from fMRI data and we will explore if community structures of these networks is different for control subjects (without schizophrenia) compared to patients diagnosed with schizophrenia. We will also explore whether community structure changes when each of the subject groups are treated with different antipsychotic drugs.

Another scenario where graphs and networks play a role in the description of time series is referred to as the "network based time series analysis" [1]. One can investigate the structure of a time series in graph-theoretic or combinatoric terms, with in principle two different objectives in mind: (i) a more *theoretical* one, where the aim is to provide a combinatorial description of certain classes of dynamics and where the objective is to produce rigorous and analytical results linking dynamical and combinatoric properties, and (ii) a more *applied* one, where the time series are considered as experimental observations, and the aim is to use this mapping as a tool for extracting (graph-theoretic) features describing each experimental signal, and to eventually feed machine learning algorithms for statistical learning purposes.

In the last ten years or so, a range of different methods have been proposed, all of them

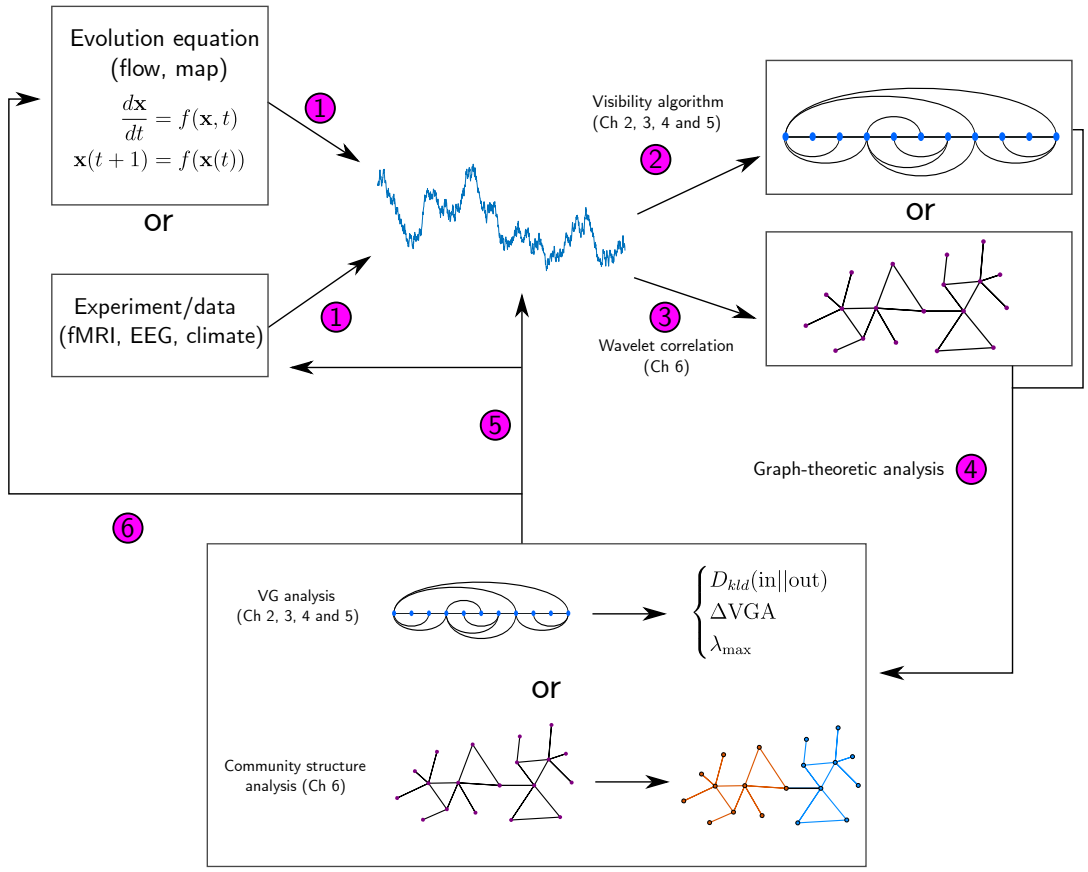


Figure 1.1: An illustration of the big picture of this thesis. We start with step 1, generating a time series from either an evolution equation, or data. We take this time series and create a graph, either with the use of the visibility algorithm (step 2) or wavelet correlation (step 3). We then perform a graph-theoretic analysis in step 4; for VG analysis we show some example quantifiers we will use, and for the community structure analysis we show a sample network with two possible communities. We then link back to the original evolution equation or data.

having a similar aim [1]. We will use two methods. First, we will use the visibility algorithm: from a given signal $x(t)$ one extracts a graph $G = (V, E)$, in such a way that the topology of G inherits or encodes the structure of the time series (and, thus, the dynamical properties of the system that generated the signal). Second, we will construct correlation networks from multiple time series (each corresponding to a different RoI), using wavelet correlation. An illustration of the big picture is provided in Figure 1.1. The first step is to extract a time series. This can be, for example, a particular trajectory of a discrete dynamical system (or a sampled sequence of a continuous flow), or from a

real world experiment (such as the amount of blood flow at a region of interest, in the brain, over time). Steps 2 and 3 in Figure 1.1 consist of mapping the information stored in the time series into a graph (where step 2 involves using the visibility graph, and step 3 is using wavelet correlation) . The crucial factor is that the mapping recipe is such that the "relevant" information stored in the time series is inherited in the structure of the network, and at the same time, the "irrelevant" information is removed. Step number 4 in Figure 1.1 consists of making a graph-theoretic analysis of the resulting structure. This process is carried out by using the tools and measures of graph theory and network science to describe the structure of the graph. We will use graph properties such as the degree sequence, the degree distribution and the spectral properties of the graph. We will then link some dynamical property of the dynamical system to a topological property of the resulting graph (step 6 in Figure 1.1), as to complete the theoretical investigation. For instance, if the dynamical process is chaotic, one can aim at linking the Lyapunov exponent (a dynamical invariant) with a topological property of the resulting graph. On the other hand, if the aim is to make use of this tool for statistical learning purposes, then we can use the set of graph-theoretic descriptors as features describing the experimental signal, which can later be used in machine learning applications (step 5 in the figure). For instance, one can aim to classify whether different subjects are healthy, or have a certain condition, by projecting the signals in a space spanned by the graph theoretic features and subsequently training a classification algorithm in this space.

The visibility graph will make up the bulk of research in this thesis (more specifically, in Chapters 2 to 5), and it deserves a brief overview. The method is inspired by the concept of visibility [2] and proceeds by mapping a time series of n data $\{x_1, x_2, \ldots, x_n\}$ into an undirected graph of n nodes according to a specific geometric criterion. As already discussed, the motivation is to subsequently make use of complex networks techniques to characterise time series. There are two types of visibility graphs: the natural visibility graph (VG) [3] and the horizontal visibility graph (HVG) [4]. In both cases, we start by constructing a vertex set V with the same cardinality of the time series n, with $|V| = n$.

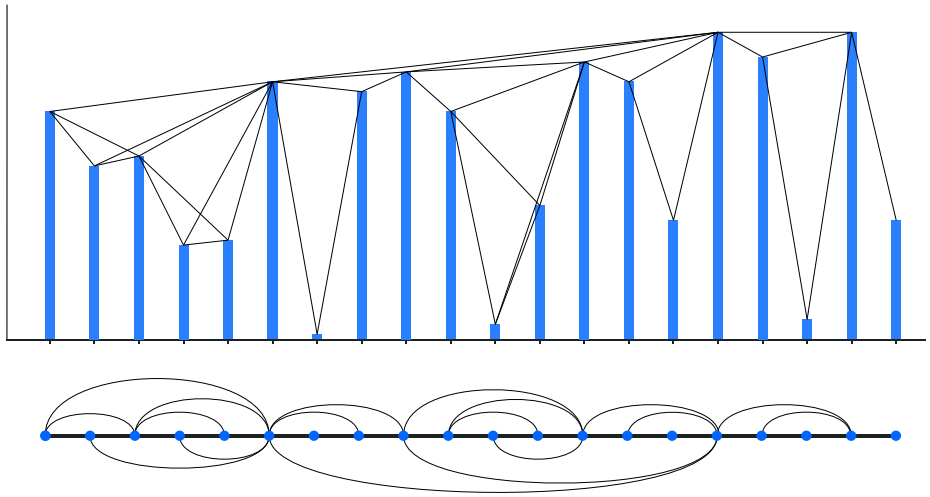


Figure 1.2: Time series (top) along with its natural visibility graph (bottom). Immediately visible is the outerplanar structure and the Hamiltonian cycle

Nodes are labelled in correspondence to the time stamp of the data, i.e., node 1 is related to datum x_1, node 2 is related to datum x_2, and so on. However, the way nodes are linked differs between the natural and the horizontal versions of the visibility graph. In the case of the natural visibility graph, any two nodes i and j are connected if the following convexity criterion is fulfilled in the data:

$$x_k < x_i + \frac{k-i}{j-i}[x_j - x_i],\ \forall k : i < k < j$$

Similarly, the horizontal visibility graph is a subgraph of the natural visibility graph, and is obtained by applying a similar procedure but with a slightly different, but stricter, linking criterion which instead only relies on the ordering of the data:

$$x_k < \inf\{x_i, x_j\},\ \forall k : i < k < j.$$

It is far easier to conceptualise the algorithm if we look at a diagram, and diagrams for the natural visibility graph and horizontal visibility graph are shown in Figure 1.2 and Figure 1.3. A more thorough definition and discussion will be introduced in Chapter 2.

In this thesis we are going to extend the visibility graph methods to deal with various particular scenarios, and apply these new extensions in several applications in finance,

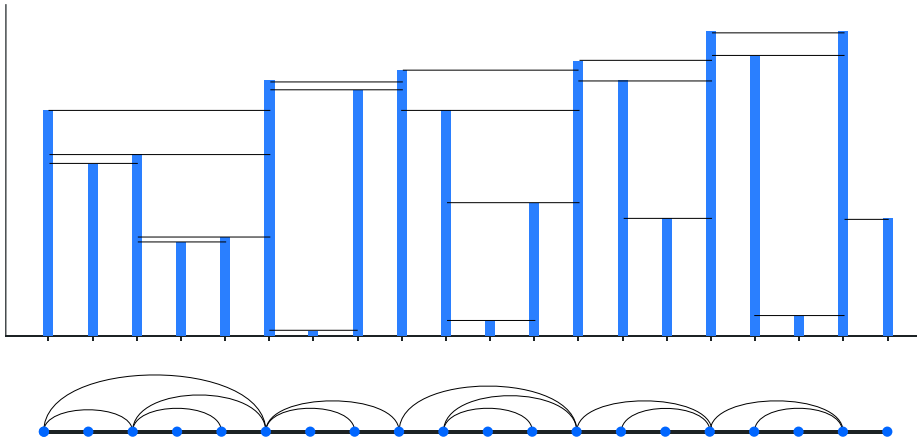


Figure 1.3: Time series (top) along with its horizontal visibility graph (bottom). This diagram is based upon the same time series as in Figure 1.2 and is therefore a subgraph of the natural visibility graph.

neuroscience and climate. We will also investigate purely theoretical questions.

The thesis is subsequently subdivided in a number of chapters. Each chapter aims to be self-contained, and corresponds to a different publication. Chapter 2 makes a theoretical investigation on how both the natural and horizontal visibility graphs work in the problem of detecting statistical time irreversibility, in the specific case where the dynamical process is non-stationary. The content of this chapter has been published in *Physical Review E* [5]. Chapter 3 makes use of the theoretical developments of chapter 2 and applies these techniques to assess the degree of time irreversibility of financial time series. The content of this chapter has been published in *Physics Letters A* [6]. In Chapter 4 we extend the two version of the visibility graphs to assess the differences between maxima and minima in time series. This is carried out by considering the time series $x(t)$, and its inverse $-x(t)$, and analysing the difference between the degree distributions of each series' visibility graphs. After a theoretical analysis of the methods, we showcase their performance in real-world data. The content of this chapter has been published in *Scientific Reports* [7]. In Chapter 5 we focus on a specific property of horizontal visibility graphs; their spectral properties. We make use of the logistic map (which generates both regular and chaotic trajectories) and make a systematic investigation of the spectral properties of the resulting graphs. The content of this

chapter has been published in *Journal of Physics A* [8]. Finally, in Chapter 6, we instead consider functional networks extracted from fMRI time series, and explore how these networks change when the subjects are treated with different antipsychotic drugs. The content of this chapter has been published in *Journal of Complex Networks* [9]. In Chapter 7 we provide a brief conclusion and discussion.

Chapter 2

Time reversibility from visibility graphs of non-stationary processes

2.1 Introduction

A dynamical process is stationary if its joint distribution does not change under time shift, hence sample time series extracted from the same process at different times have similar statistics, with small deviations only occurring as finite size effects. A stationary time series $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ is called statistically time reversible if the series and its time reversed $\mathcal{S}^* = \{x_n, x_{n-1}, \ldots, x_1\}$ are equally likely, i.e., if they have identical joint distributions [10].

For instance, Gaussian linear processes such as white noise, or conservative chaotic processes such as Hamiltonian chaos are time reversible, and related to processes in thermodynamic equilibrium in statistical physics. Non-linear stochastic processes, or dissipative chaotic processes are generally found to be irreversible [11], and are associated to processes that operate away from equilibrium in a thermodynamic sense. For these cases,

recent works relate the amount of entropy that a system is producing while being away from equilibrium to the amount of time irreversibility, computed from the time evolution of adequate physical observables [12].

Traditionally, the study of statistical time irreversibility has only applied to stationary processes [10]. A dynamical process is stationary if its joint distribution does not change under time shift, hence sample time series extracted from the same process at different times have similar statistics, with small deviations only occurring as finite size effects. For these processes, one can then meaningfully estimate properties about the underlying stationary distribution of the process (if this exists) through its estimation for finite series. In particular, one can quantify the amount of time irreversibility in stationary processes via a number of strategies and algorithms proposed in the literature, including simple statistical differences between forward and backward trajectories [11–15] or more sophisticated methods such as compression [16]. In every case, note that time series need to be symbolized before an irreversibility measure can be computed [11]. Via fluctuation theorems, a remarkable identity between the Kullback-Leibler divergence of the forward and backward statistics of a time series (i.e., the statistics of *individual particle trajectories*) and the amount of entropy that the underlying thermodynamic system is producing has been found recently [12, 13], which has further stimulated the study of time series irreversibility in statistical physics. On the other hand, non-stationary processes have underlying joint distributions that change over time, hence no straightforward quantification of the time asymmetry of a process can be extracted from the analysis of finite series. The precise definition of time series reversibility in time series analysis is the following: a time series $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ is called statistically time reversible if the time series $\mathcal{S}^- = \{x_{-1}, x_{-2}, \ldots, x_{-n}\}$ has the same joint distribution as $\mathcal{S}$ [17]. By definition, non-stationary series are (infinitely) irreversible: the statistical properties of a non-stationary process vary with time, and therefore $\mathcal{S}$ and $\mathcal{S}^-$ have different statistics that increase over time without bounds. It is only for stationary processes where the standard definition of time reversibility acquires its full meaning. For this latter case,

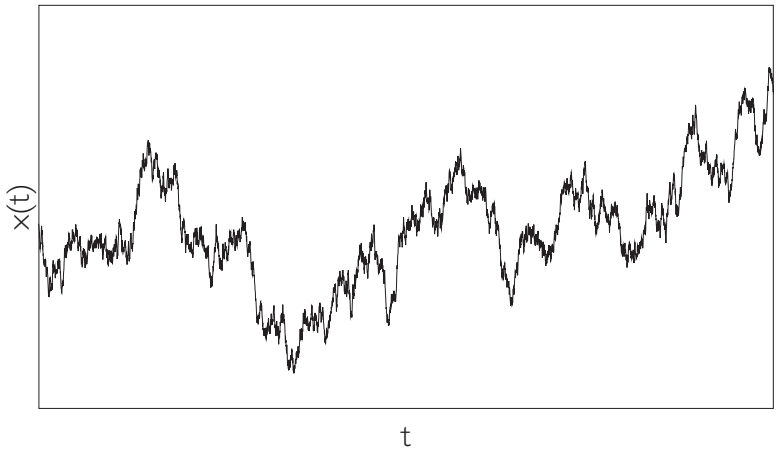


Figure 2.1: Sample time series of an unbiased random walk, as canonical example of a non-stationary process. By definition, the process is (infinitely) time irreversible, although if we remove the Y axis, then it is impossible to know in which direction time is flowing, as both pictures (forward and backward) are equally likely.

$\{x_{-1}, x_{-2}, \ldots, x_{-n}\}$ and $\{x_{-1+m}, x_{-2+m}, \ldots, x_{-n+m}\}$ have the same joint distributions $\forall m$, so for the particular choice $m = n + 1$, the definition of time reversibility reduces to the equivalence between forward and backward statistics. Hence the popular motto "time reversibility implies stationarity" [10]. Note however, that if we understand the source of irreversibility in close relation to directionality (or, in other words, underlying sources of memory), then one could argue that there should exist different degrees of irreversibility in non-stationary processes: for instance, a Markovian random walk should arguably be "less irreversible" than a non-Markovian one, even if both are non-stationary. To further illustrate this, in Figure 2.1 we plot a realisation of a 1d random walk $x(t)$ that starts at the origin $x(0) = 0$, where we have deliberately removed the vertical axis. While this is a non-stationary process and hence time irreversible, could the one assert which is the correct direction of time? Wouldn't both the forward and backward processes be equally likely, once the vertical axis is removed? This figure could indeed both plot $\{x_1 \ldots x_n\}$ or $\{x_{-1} \ldots x_{-n}\}$. Moreover, if $x(t)$ describes the trajectory of a Brownian particle in a system in thermal equilibrium, shouldn't this time series on average have a null entropy production - hence a reversible character?

In this chapter we show that, by suitably mapping non-stationary time series into a graph-theoretical setting by means of the visibility algorithm (which we will formally define), we can actually quantify different kinds of time asymmetries in the underlying dynamics on non-stationary processes, where random walks such as the one presented in Figure 2.1 are indeed time reversible in the new framework. The family of visibility algorithms were recently introduced as simple mappings between time series and graphs, with the aim of enabling the description and classification of the structure of time series as well as their underlying dynamics in graph-theoretic terms. Among other interesting advantages, these methods do not require the series to be pre-symbolized. In the context of time series irreversibility, a directed version of visibility algorithms was also proposed recently to assess irreversibility in stationary real-valued time series [18, 19], and has been used extensively [20–23]. Here we extend that former analysis to the realm of non-stationary signals. We investigate the topological properties of the visibility graphs (VGs) and horizontal visibility graphs (HVGs) associated to several types of non-stationary processes, and pay particular attention to their performance in quantifying several degrees of irreversibility. We take advantage from the fact that the topological properties of these graphs are effectively invariant under time shift for large classes of non-stationary processes, which allows us to introduce the concept of visibility graph stationarity. This in turn allows us to compare to extract meaningful information on the time asymmetry of non-stationary processes.

The rest of the chapter is as follows: in Section 2.2, we recall how univariate real-valued time series can be mapped into the family of visibility graphs (natural and horizontal versions), and explain how a directed version of these graphs can be used to estimate statistical time irreversibility of the original time series, without requiring to symbolise the series. We summarize previous findings on canonical stationary processes and prove a lemma that permits us to quantify the degree of irreversibility in non-stationary ones. In Section 2.3, we focus on random additive processes, and provide some exact results on the properties of visibility graphs associated to simple random walks. We prove that

unbiased random walks are indeed time reversible according to new definitions, and that for biased ones, the HVG method can quantify the degree of irreversibility. In Section 2.4, we extend these results to random multiplicative processes. We numerically explore the performance of visibility methods in these cases and complement these findings with some analytical and heuristic explanations. In Section 2.5 we conclude.

2.2 Measuring irreversibility using visibility graphs

Here we first introduce a formal definition of the visibility and horizontal visibility graphs associated to an ordered series of real-valued data. These are inspired in computational geometry [2] and the intuition underlying the mappings (in particular, the link criteria) shares some similarities with first passage time statistics [24]. We also introduce the notions of VG (HVG)-stationarity and VG (HVG)-irreversibility, which we will rely on subsequently.

Definition 2.2.1 (Natural Visibility Graph (VG))**.** Let $\mathcal{S} = \{x_1, \ldots, x_n\}$ be a real-valued scalar time (or otherwise ordered) series of n data. A visibility graph (VG) is an undirected graph of n nodes, where each node $i \in [1, n]$ is labelled by the time order of its corresponding datum x_i. Hence x_1 is mapped into node $i = 1$, x_2 into node $i = 2$, and so on. Then, two nodes i, j (assume $i < j$ without loss of generality) are connected by a link if and only if one can draw a straight line connecting x_i and x_j that does not intersect any intermediate datum x_k, $i < k < j$. Equivalently, i and j are connected if the following *convexity* criterion is fulfilled:

$$x_k < x_i + \frac{k-i}{j-i}[x_j - x_i], \ \forall k : i < k < j.$$

See Figure 1.2 in Chapter 1, for a sample time series with its respective visibility graph. By construction, VGs are planar, connected graphs, and this construction is invariant under a set of basic transformations in the series, including horizontal and vertical translations. An important property of these graphs is that they are well suited to investigate

the properties of non-stationary signals. For instance, it was shown [25] that the degree distribution of VGs associated to series generated by a (non-stationary) fractional Brownian motion with Hurst exponent H have a power-law tail with exponent $\gamma = 3 - 2H$. This analysis has been subsequently applied to finance [26, 27], fluid dynamics [28, 29], and medical research [30], to cite a few. Other works applying VG to finance deal with properties such as spanning trees [31], or community structure [32].

The horizontal visibility graph is a subgraph of the natural visibility graph, and is obtained by applying a similar procedure but with a slightly different, but stricter, linking criterion which instead only relies on the ordering of the data:

Definition 2.2.2 (Horizontal Visibility Graph (HVG))**.** Let $\mathcal{S} = \{x_1, \ldots, x_n\}$ be a real-valued scalar time (or otherwise ordered) series of n data. A horizontal visibility graph (HVG) is an undirected graph of n nodes, where each node $i \in [1, n]$ is labelled by the time order of its corresponding datum x_i. Hence x_1 is mapped into node $i = 1$, x_2 into node $i = 2$, and so on. Then, two nodes i, j (assume $i < j$ without loss of generality) are connected by a link if and only if one can draw a *horizontal* line connecting x_i and x_j that does not intersect any intermediate datum x_k, $i < k < j$. Equivalently, i and j are connected if the following *ordering* criterion is fulfilled:

$$x_k < \inf(x_i, x_j), \ \forall k : i < k < j$$

See Figure 1.3, in Chapter 1, for a sample time series with its respective horizontal visibility graph.

Definition 2.2.3 (VG stationarity)**.** A dynamical process $\{X_t\}$ is said to be VG-stationary if and only if the topological properties of the VG associated to a sample time series of size n extracted from $\{X_t\}$ are asymptotically (i.e., for large n) invariant under time shift (in the statistical sense). In other words, processes for which sample time series $\{x_1, x_2, \ldots, x_n\}$ and $\{x_{1+\tau}, x_{2+\tau}, \ldots, x_{n+\tau}\}$, for all τ, generate (in the limit

of large n) statistically equivalent VG, are called VG-stationary. In particular, the degree distributions of VG associated to $\{x_1, x_2, \ldots, x_n\}$ and $\{x_{1+\tau}, x_{2+\tau}, \ldots, x_{n+\tau}\}$ are asymptotically (for large n) identical for VG stationary processes.

Similarly, we define HVG stationarity in the same way.

Theorem 2.2.4. *Let $\{X_t\}$ be a non-stationary process, and consider two time series samples of n data extracted from $\{X_t\}$: $\{x_1, x_2, \ldots, x_n\}$ and $\{x_{1+\tau}, x_{2+\tau}, \ldots, x_{n+\tau}\}$ for some $\tau \in \mathbb{Z}$. If $\forall\tau\ \exists c \in \mathbb{R}$ such that $\{x_1, x_2, \ldots, x_n\}$ and $\{x_{1+\tau}+c, x_{2+\tau}+c, \ldots, x_{n+\tau}+c\}$ are statistically equivalent time series (i.e., have the same joint distributions), then the process $\{X_t\}$ is VG (HVG) stationary.*

Proof. Both VG and HVG are invariant under vertical rescaling of the time series [3], that is to say, the series $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{S}' = \{x_1 + c, x_2 + c, \ldots, x_n + c\}$ generate the same VG and HVG $\forall c \in \mathbb{R}$. Thus $\{x_{1+\tau} + c, x_{2+\tau} + c, \ldots, x_{n+\tau} + c\}$ and $\{x_{1+\tau}, x_{2+\tau}, \ldots, x_{n+\tau}\}$ also generate the same VG and HVG, $\forall c \in \mathbb{R}$. Choose c such that $\{x_1, x_2, \ldots, x_n\}$ and $\{x_{1+\tau} + c, x_{2+\tau} + c, \ldots, x_{n+\tau} + c\}$ are statistically equivalent (i.e., they have identical asymptotic joint distributions). Note that in additive processes, this usually is fulfilled for $c = x_1 - x_{1+\tau}$, but should be set for each process independently. Then, $\{x_{1+\tau}+c, x_{2+\tau}+c, \ldots, x_{n+\tau}+c\}$ and $\{x_1, x_2, \ldots, x_n\}$ will also generate statistically equivalent VG and HVG, hence by definition the process is VG and HVG stationary. □

Both the VG and HVG have been fruitfully applied in recent years to describe and classify different types of time series and dynamics. For instance, VG have been shown to be a viable method to quantify the Hurst exponent of fractional Brownian motion (inherently non-stationary signals); a linear relation was found between the Hurst exponent H of a time series and the exponent γ of the power law degree distribution of the associated VG, $\gamma = 3 - 2H$ [25]. The HVG has been used in turn to describe chaotic and correlated stochastic processes [33], or to provide a graph-theoretical description of canonical routes to chaos [26, 34, 35], and it has been shown that HVGs are analytically tractable for

several classes of Markovian dynamics [36]. Both VG and HVG are connected planar graphs by construction, which have a Hamiltonian path described by the path $1-2-\cdots-n$. HVG are outerplanar graphs, and again by construction, one can easily prove that the HVG of $\mathcal{S}$ is a subgraph of VG. As both VG and HVG have a natural order induced by the time arrow (or equivalently, by the order of the associated series $\mathcal{S}$), it is natural to define the degree sequence of a VG or a HVG as $\{k(t)\}_{t=1}^{n}$, where $k(t)$ is the degree of node $i=t$.

Note that previous definitions generate undirected graphs. However, these can be made *directed* by again assigning to the links a time arrow. Accordingly, a link between i and j (where time ordering yields $i<j$), generates an *outgoing* link for i and an *ingoing* link for j. The degree sequence thus splits into an ingoing degree sequence $\{k_{in}(t)\}_{t=1}^{n}$, where $k_{in}(t)$ is the ingoing degree of node $i=t$, and an outgoing degree sequence. An important property at this point is that the ingoing and outgoing degree sequences are interchangeable under time series reversal. That is to say, if we define the time reversed series $\mathcal{S}^{*}=\{x_{n+1-t}\}_{t=1}^{n}$, then we have the following identities

$$\{k_{in}(t)\}[\mathcal{S}]=\{k_{out}(t)\}[\mathcal{S}^{*}];\ \{k_{out}(t)\}[\mathcal{S}]=\{k_{in}(t)\}[\mathcal{S}^{*}] \tag{2.1}$$

Now, one can define, from the ingoing and outgoing degree sequences, an ingoing degree distribution $P(k_{in})\equiv P_{in}(k)$ and an outgoing degree distribution $P(k_{out})\equiv P_{out}(k)$, and property (Equation (2.1)) is inherited in the distributions, such that

$$P_{in}(k)[\mathcal{S}]=P_{out}(k)[\mathcal{S}^{*}];\ P_{out}(k)[\mathcal{S}]=P_{in}(k)[\mathcal{S}^{*}] \tag{2.2}$$

Definition 2.2.5 (VG and HVG reversibility)**.** In this thesis, a time series $\mathcal{S}=\{x(t)\}_{t=1}^{n}$ is said to be (order p) VG-reversible (HVG-reversible) if and only if, for large n, the order p block *in* and *out* degree distribution estimates of the VG (HVG) associated to $\mathcal{S}$ are asymptotically identical:

$$P_{in}(k_1k_2\ldots k_p)=P_{out}(k_1k_2\ldots k_p).$$

Remark. According to property (Equation (2.2)), Definition 2.2.5 implies that, under the VG/HVG setting, the statistics of the degree sequences are statistically invariant under time reversal.

Other topological properties of VG/HVG could be used to quantify time asymmetries, as has been reported recently [19]. For finite series, we will assess how close the system is to reversibility by quantifying the distance (in distributional sense) between P_{in} and P_{out}. While several possible measures can be used, here we focus on the Kullback-Leibler divergence between the *in* and *out* distributions, previously proposed in [18]:

Definition 2.2.6 (Kullback-Leibler divergence or KLD)**.**

$$D_{kld}(\text{in}||\text{out}) = \sum_k P_{in}(k) \log\left(\frac{P_{in}(k)}{P_{out}(k)}\right)$$

$D_{kld}(in||out)$ is a semi-distance which is null if and only if $P_{in}(k) = P_{out}(k)$, and is positive otherwise. As a technical remark, note that $D_{\text{kld}}(P_{in}|P_{out})$ diverges if p and q have different supports (i.e, if $q(m) = 0, p(m) \neq 0$ or $p(m) = 0,\ q(m) \neq 0$ for some value m). In order to appropriately weight this possibility while maintaining a finite irreversibility measure, a common procedure is to introduce a small bias that allows for the possibility of having a small uncertainty for every contribution [13]. Here we introduce a bias of order $\mathcal{O}(1/n^2)$ where n is the series size (i.e. we replace all vanishing frequencies with $1/n$, and we normalize the frequency histogram appropriately). We then redefine VG/HVG-reversibility as

$$\lim_{n\to\infty} D_{kld}(\text{in}||\text{out}) = 0$$

Truly irreversible processes will have positive irreversibility values even in the limit of large n: we will call these processes VG/HVG-irreversible. For VG/HVG-reversible processes, $D_{kld}(in||out)$ will have a positive finite value for finite size series that approaches 0 as size increases, with $D_{kld}(in||out) \propto n^{-\delta}$ (where δ will be different for VG and HVG). As the convergence is relatively slow, the finite-size results will also be helpful to com-

pare and classify the degrees of reversibility of finite series across different processes, something relevant in practice.

We have checked that all the results we found in this chapter are qualitatively equivalent under alternative distance measures between distributions, such as the Manhattan (L_1) distance $D_{L1} = \sum_k |P_{in}(k) - P_{out}(k)|$, although in this latter case, convergence to zero for reversible cases is typically slower (results not shown). We also chose the Kullback-Leibler divergence one as it has some physical meaning: for stationary series, $D_{kld}(in||out)$ provides a lower bound [18] to the thermodynamic entropy that a non-equilibrium steady state described by a state variable $x(t)$ is producing along its time evolution. Also, as *in* degrees account for past information while *out* degrees account for future information (or past information in the time reversed case), then D_{kld}(in||out) is formally akin to D_{kld}(forward||backwards) in graph space, whereas or D_{kld}(out||in) is the formal analogue to D_{kld}(backwards||forward). This measure was used to assess HVG-reversibility in the context of stationary processes and non-equilibrium steady states. Here we further extend that analysis to investigate both HVG and VG reversibility for several classes of dynamics. In what follows, we drop the specification and in the text we refer to D_{kld}(in||out) $\equiv D$.

In this chapter, we will mainly look at $p = 1$, so for readability we will drop this specification. Some important remarks are in order. First, note that there is no direct equivalence between order p VG (HVG) reversibility in stationary processes and order p reversibility in the time series, expressed as $P(x_1, \ldots, x_p) = P(x_p, \ldots, x_1)$. As a matter of fact, the degree of each node in a VG(HVG) graph inherits information from the whole time series, hence it is a global measure. Nonetheless, as we only look at order $p = 1$, we can't rule out the possibility that certain processes appear to be VG/HVG reversible at order $p = 1$ but are found to be irreversible at higher orders, as happens for time series produced out of equilibrium where the net current is balanced to zero via stalling forces [18]. So whereas in this chapter we are dropping the "order p" for readability, the reader should recall that we are working at order $p = 1$ in the VG/HVG setting.

Second, it is important to highlight that standard methods that aim to quantify time series reversibility usually address the statistical differences of time series directly. As already stated, the original definition of time reversibility precludes the possibility of quantifying irreversibility in non-stationary signals: $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{S}^- = \{x_{-1}, x_{-2}, \ldots, x_{-n}\}$ have statistical differences that grow with n for non-stationary processes.

Third, in order to assess irreversibility directly in real-valued data, one is unavoidably required to *symbolize* the series in advance: one needs to pre-define an alphabet (with an arbitrary number of symbols) and generate a time series partition to map each datum into a symbol. Both the alphabet and the partition have to be defined ad hoc, and results often depend on these free parameters, which inevitably generates ambiguities in finite size. Furthermore, in the non-stationary realm, symbolisation is clearly ill-defined as the phase space itself grows with the series size.

Here, we take advantage of the properties of the visibility algorithms, and apply the irreversibility measures directly on the degree sequences $\{k(t)\}_{t=1}^{T}$, where $k(i)$ is the degree of node i. This sequence is discrete by construction, so there is no need to perform any ad hoc symbolisation.

2.2.1 Preamble on stationary systems: white noise versus fully developed chaos

As an illustration, let us begin by considering two paradigmatic stationary processes. The first one is white noise, a stationary and statistically time reversible uncorrelated stochastic process. Consider a sequence of i.i.d. uncorrelated random variables (i.e., $\langle \xi(t)\xi(t+\tau)\rangle = \delta(\tau)$) extracted from some probability distribution $p(x)$ with some compact real support as a realization of the white noise process. For this process, a theorem [18] guarantees that, asymptotically, $P_{in}(k) = P_{out}(k) = 2^{-k}$ and as a consequence, the process is HVG-reversible $\forall p(x)$. As we lack equivalent theorems for VGs, we have run numerical simulations. In Figure 2.2 we plot, in semi-log, $P_{in}(k)$ and $P_{out}(k)$ for the VG

associated to a sample of 2^{15} i.i.d. uniform random variables $\sim U[0,1]$. In panel (b) of the same figure we plot the irreversibility estimate D for increasing system size (each dot is an average over 10 realizations). We conclude that white noise is both HVG and VG-reversible showing that this process is indeed VG-reversible, in good agreement with previous theory.

For comparison, we also consider the fully chaotic logistic map $x_{t+1} = 4x_t(1-x_t)$, where $x \in [0,1]$, a paradigmatic deterministic stationary process which is nonetheless time irreversible. HVG-irreversibility of the fully chaotic logistic map was shown in [18], where it was found that $P_{in}(k)$ and $P_{out}(k)$ were asymptotically different distributions. We can summarize this by computing $P_{in}(1)$ and $P_{out}(1)$, and showing that they are strictly different. We first rely on the fact that this map is Markovian, hence

$$P_{out}(k=1) = \int_0^1 dx_t \int_{x_t}^1 dx_{t+1} f(x_t) f(x_{t+1}|x_t),$$
$$P_{in}(k=1) = \int_0^1 dx_t \int_{x_t}^1 dx_{t-1} f(x_{t-1}) f(x_t|x_{t-1}).$$

where $f(x)$ is the invariant probability measure that characterizes the long-term fraction of time spent by the system in the various regions of the attractor. In the case of the (fully chaotic) logistic map the attractor is the whole interval $[0,1]$ and the invariant measure is

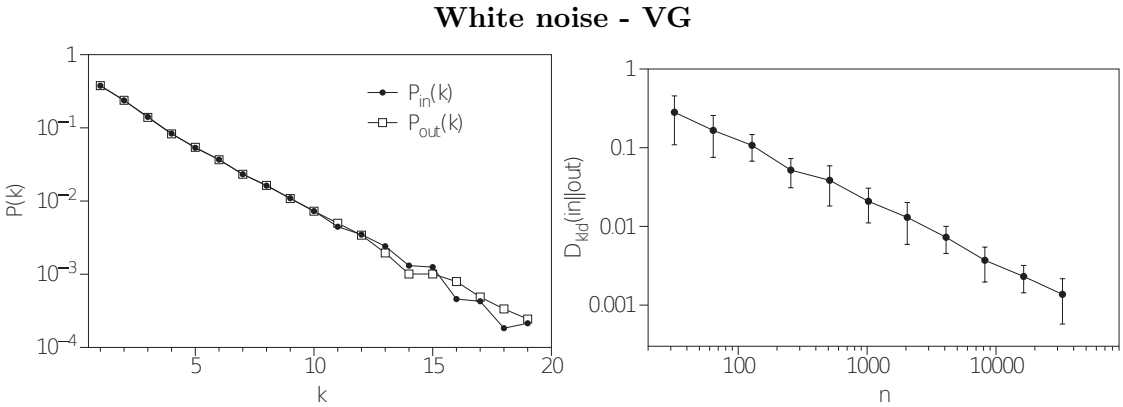
$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}. \tag{2.3}$$

Now, for a deterministic system, the transition probability is simply

$$f(x_{t+1}|x_t) = \delta(x_{t+1} - F(x_t)),$$

where $\delta(x)$ is the Dirac delta distribution and $F(x) = 4x(1-x)$. Notice that, using the properties of the Dirac delta distribution, $\int_{x_t}^1 \delta(x_{t+1} - F(x_t))dx_{t+1}$ is equal to one if and only if $F(x_t) \in [x_t, 1]$, which happens for $0 < x_t < 3/4$, and it is zero otherwise. Therefore the only effect of this integral is to restrict the integration range of x_t to be $[0, 3/4]$:

$$P_{out}(k=1) = \int_0^{3/4} dx_t f(x_t) = 2/3.$$



(a) (b)

Figure 2.2: (a) Semi-log plot of the in and out degree distributions of the natural visibility graph associated to a time series of 2^{15} i.i.d. uniformly random uncorrelated variables $\sim U[0,1]$. Both distributions are identical up to finite-size effects fluctuations, suggesting that the underlying process is VG-reversible. (b) Log-log plot of the irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 10 realizations). This measure vanishes asymptotically as $1/n$, showing that finite irreversibility values for finite size are due to statistical fluctuations that vanish asymptotically.

Similarly,

$$P_{in}(k=1) = \int_{3/4}^{1} f(x_t)dx_t = 1/3.$$

We conclude that $P_{out}(1) \neq P_{in}(1)$ for the fully chaotic logistic map. Since D is semi-positive definite and null if and only if the two distributions are identical, then D is strictly positive for this process, i.e., it is HVG-irreversible.

Because we don't have equivalent theory for VGs, we have again run numerical simulations for this case, which are plotted in Figure 2.3. Once again, the *in* and *out* distributions are clearly different and their Kullback-Leibler divergence converges to a finite, positive value as the series size increases, also suggesting VG-irreversibility.

In what follows we extend previous studies on stationary signals to the realm of non-stationary time series.

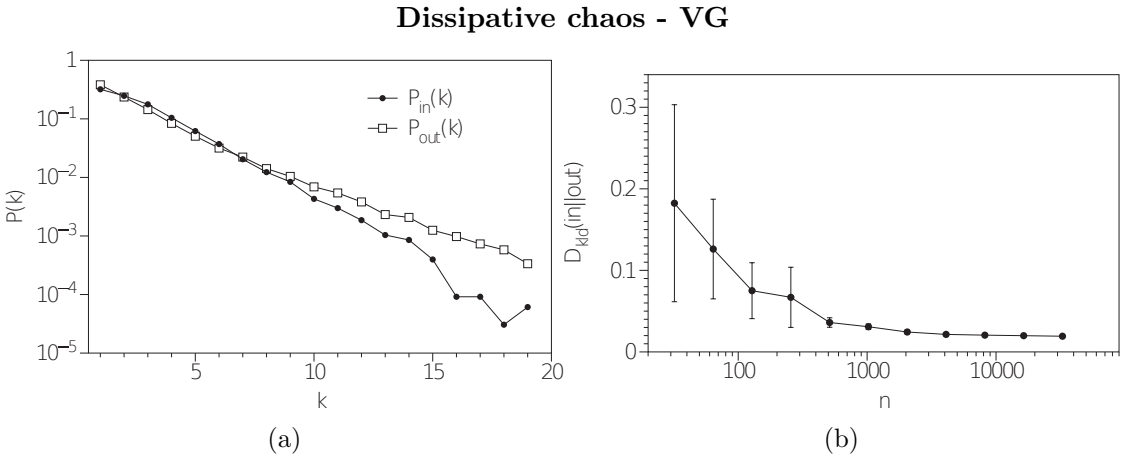


Figure 2.3: (a) Semi-log plot of the in and out degree distributions of the natural visibility graph associated to a time series of 2^{15} data generated from a fully chaotic logistic map $x(t+1) = 4x(t)(1-x(t))$. Distributions are clearly different, suggesting that the underlying process, although stationary, is VG-irreversible. (b) Irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 10 realizations). This measure converges to a finite value with series size, confirming that the process yields a positive irreversibility measure.

2.3 Additive random walks

2.3.1 Simple random walks

Let us start by considering a simple one dimensional random walk, described by

$$x(t+1) = x(t) + \xi, \ \xi \in \{-1, 1\}, \tag{2.4}$$

i.e., the step distribution is the Rademacher-1/2 distribution. Without loss of generality, if we generate a time series of n data $\{x_1, \ldots, x_n\}$ which deterministically starts in the origin (for which $\mathbb{E}(x_1) = \sigma^2(x_1) = 0$), as the process is unbiased, we have $\mathbb{E}(x_n) = 0 \ \forall n$, but, in virtue of the central limit theorem, the variance fulfils $\sigma^2(x_n) \sim n$, so the process is non-stationary. In this subsection we derive exact results on the in/out degree distributions for this simple process.

Theorem 2.3.1. *The in and out degree distributions of the HVG associated to a bi-*

infinite series generated by a 1d simple random walk are

$$P_{in}(k) = P_{out}(k) = \begin{cases} 1/2 & k = 1, 2 \\ 0 & otherwise \end{cases} \tag{2.5}$$

Proof. First, notice that we don't necessarily need to compute $P_{out}(k)$ and $P_{in}(k)$ separately. We use property 2.5 and focus on both P_{out} as applied to the time series and its time reverse, that we label P_{out} and P^*_{out} respectively.

- $k = 0$: by construction there is exactly one node with $k_{out} = 0$ (the final node) and only one node with $k_{in} = 0$ (the initial node), so $P_{out}(0) = P^*_{out}(0) = 1/N$, where N is the series size. Hence for bi-infinite series, $P_{out}(0) = P^*_{out}(0) = 0$.

- $k = 1$: $P_{out}(1) = \text{prob}(x_{t+1} \geq x_t) = 1/2$, $P^*_{out}(1) = \text{prob}(x_t \geq x_{t+1}) = 1/2$.

- $k > 2$: let us prove by contradiction that $P_{out}(k > 2) = P^*_{out}(k > 2) = 0$. In order for $P_{out}(k > 2) > 0$, there should be at least an ordering of data which allows that a node chosen at random has degree $k_{out} > 2$. Let us assume that the node associated to datum x_0 is that node, which *at least* has *out* visibility of the node associated to x_1 (by construction), x_p (for some $1 < p < q$) and x_q (for some $q > 2$). The geometrical restrictions on the data that follow from the horizontal visibility criterion are $\{x_0 > x_1;\ x_0 > x_p > x_1,\ x_q > x_p\}$. The first restriction yields $x_1 = x_0 - 1$ according to Equation (2.4). On the second condition we have $x_p > x_1$ that implies $x_p \geq x_0$. But this contradicts the first inequality of the second restriction, $x_0 > x_p$. Hence $P_{out}(k > 2) = 0$. A similar geometrical argument yields $P^*_{out}(k > 2) = 0$.

- Normalization of the probability yields $P_{out}(2) = P^*_{out}(2) = 1/2$, which concludes the proof.

□

Remark. In the previous proof, we didn't need to explicitly compute $P_{out}(2)$. As a

curiosity, we will show that this is possible using simple enumerative combinatoric arguments. We start by using the diagrammatic approach proposed in [36], which divides the computation of each degree probability into an infinite sum of corrections of order α, $P_{out}(2) = \sum_{\alpha=0}^{\infty} P_{out}^{(\alpha)}(2)$, where α is the number of hidden variables (hidden data) in a given configuration. That is to say, $P_{out}^{(\alpha)}(2)$ gathers the contribution given by all the diagrams for which we find $k_{out} = 2$, that include a total of α hidden variables (hidden data with no visibility). For instance, for $k_{out} = 2$ there is exactly one path (diagram) at order $\alpha = 0$, that can be labelled as $\{BT\}$, where B stands for a movement downhill ($\xi = -1$) and T stands for a single movement uphill ($\xi = +1$). This represents the diagram $\{x_0, x_1, x_2\}$ where $x_1 = x_0 - 1,\ x_2 = x_0$, and its associated probability is directly 2^{-2}. There are no contributing paths at order $\alpha = 1$ (in fact, all odd values of α are forbidden by construction), whereas there is exactly one path at order $\alpha = 2$, labelled as $\{BBTT\}$, that contributes with a probability 2^{-4}. In fact, any path should start and end by $\{B|\cdots|T\}$. The number of hidden variables α is represented here as the number of extra letters to be located. While there are a total of 2^{α} possible paths that start with a downhill movement and end with an uphill movement (with equal weight $2^{-(\alpha+2)}$), not all of them are allowed in the sense of generating a valid path for $k_{out} = 2$ - only strictly negative closed walks of length $\alpha + 2$ are allowed at order α.

First, $k_{out} = 2$ requires that the initial and final node have associated data of identical height. Since the initial movement is downhill (B) and the final one is uphill (T), the hidden variables should contribute with a null vertical movement, so half of them have to be involved in a downhill movement, and half of them in an uphill one. This reduces the number of paths from 2^{α} to $\binom{\alpha}{\alpha/2}$. Furthermore, only those paths that always remain under x_0 until reaching the end datum will actually be paths of order α (if they cross the x_0 level at prior stages they are considered corrections of lower order).

Interestingly, the number of allowed paths can then be seen as the number of words of length α having $\alpha/2$ B's and $\alpha/2$ T's, such that no initial segment of the word has more T's than B's. These paths are sometimes called *Dyck words* in enumerative

combinatorics. The number of Dyck words of size α is $\mathcal{C}_{\alpha/2}$, where $\mathcal{C}_n$ is the Catalan number

$$\mathcal{C}_n = \frac{1}{n+1}\binom{2n}{n}$$

Hence $P_{out}(2)$ takes the form

$$P_{out}(2) = \sum_{\alpha=0,\text{even}}^{\infty} \mathcal{C}_{\alpha/2}\left(\frac{1}{2}\right)^{\alpha+2} = \frac{1}{4}\sum_{\gamma=0}^{\infty}\frac{1}{\gamma+1}\binom{2\gamma}{\gamma}\left(\frac{1}{4}\right)^{\gamma} \tag{2.6}$$

where we have used the change of variable $\gamma = \alpha/2$. Leaving the pre factor $1/4$ aside, (Equation (2.6)) is the generating function of the Catalan numbers evaluated at $z = 1/4$. The generating function sums up to $[1 - \sqrt{1-4z}]/2z$, thus $P_{out}(2) = 1/2$, in good agreement with previous theorem.

By virtue of Theorem 2.2.4, the process described in Equation (2.4) is VG and HVG-stationary (choosing c such that every sample time series starts, say, a the origin, makes them statistically indistinguishable). Accordingly, one is entitled to explore the time asymmetries taking place in the graph space. According to theorem 1, as both the in and out degree distributions are equivalent for the HVG, the process is indeed HVG-reversible. We will now explore a generalization of this process and the performance of both HVG and VG.

2.3.2 Unbiased additive random walks

Let us generalize the previous simple random walk by considering an unbiased additive random walk

$$x(t+1) = x(t) + \xi, \quad \langle \xi \rangle = 0 \tag{2.7}$$

where ξ are i.i.d. random variables extracted from some (arbitrary) symmetric distribution. This process is, for instance, a (1d) discrete model of a Brownian particle evolving in an infinitely large system which is in thermodynamic equilibrium, a system which on average is not producing entropy. From a time series perspective, it is however a non-stationary process, i.e. time irreversible. The following theorem uses the VG and HVG method to somehow reconcile both aspects.

Theorem 2.3.2. *A bi-infinite time series generated from the unbiased random walk model defined in Equation* (2.7) *is both VG and HVG reversible.*

Proof. The first step is to prove that the process described in Equation (2.7) is both VG and HVG stationary. Choose $c = x_1 - x_{1+\tau}$ in Theorem 2.2.4, for which

$$\begin{aligned}\{x_{1+\tau} + c, x_{2+\tau} + c, \dots, x_{n+\tau} + c\} &= \{x_1, x_1 + (x_{2+\tau} - x_{1+\tau}), \dots, x_1 + (x_{n+\tau} - x_{1+\tau})\} \\ &= \{x_1, x_1 + \xi, \dots, x_1 + \sum_{i=1}^{n-1} \xi\} \\ &= \{x_1, x_2, \dots, x_n\}.\end{aligned}$$

Therefore the process is VG and HVG stationary, concluding the first part of the proof.

Accordingly, reversibility reduces to investigate whether the VG/HVG of $\{x_1, x_2, \dots, x_n\}$ and $\{x_n, x_{n-1}, \dots, x_1\}$ are statistically identical. To address this, we recall that visibility algorithms (both VG and HVG) are invariant under vertical rescaling [3, 4]. This means that two time series $\{x_1, x_2, \dots, x_n\}$ and $\{x_1 + c, x_2 + c, \dots, x_n + c\}$ yield the same VG (and the same HVG) $\forall c \in \mathbb{R}$. In particular, the (vertically shifted) reversed time series $\{x_n + c, x_{n-1} + c \dots, x_1 + c\}$ and the reverse time series $\{x_n, x_{n-1} \dots, x_1\}$ also yield the same VG and HVG $\forall c \in \mathbb{R}$. Our strategy then consists in proving that there exists a value c for which $\{x_n + c, x_{n-1} + c \dots, x_1 + c\}$ and $\{x_1, x_2, \dots, x_n\}$ are statistically identical. Choose $c = x_1 - x_n$, for which $\{x_n + c, x_{n-1} + c, \dots, x_1 + c\} = \{x_1, x_1 - (x_n - x_{n-1}), x_1 - (\xi + x_{n-1} - x_{n-2}), \dots\} = \{x_1, x_1 - \xi, x_1 - \sum_{i=1}^{2} \xi, \dots, x_1 - \sum_{i=1}^{n-1} \xi\}$. Note that in the last series, since ξ has a symmetrical distribution for the process under study, then it is invariant under the transformation $\xi \to -\xi$. So $\{x_1, x_1 - \xi, x_1 - \sum_{i=1}^{2} \xi, \dots, x_1 - \sum_{i=1}^{n-1} \xi\}$ and $\{x_1, x_1 + \xi, x_1 + \sum_{i=1}^{2} \xi, \dots, x_1 + \sum_{i=1}^{n-1} \xi\}$ are statistically equivalent. But this latter series is equivalent by definition to $\{x_1, x_2, \dots, x_n\}$, thus concluding that the process described in Equation (2.7) is both VG and HVG reversible. □

In Figures 2.4 and 2.5, we plot the results of numerical simulations on the VG and HVG respectively, for the case of an unbiased random walk where $\xi \sim U[-0.5, 0.5]$. We

find that the *in* and *out* degree distributions of both graphs coincide (up to finite size effects), in good agreement with previous theorem. The irreversibility measure (panel b) for the HVG case decreases monotonically with series size n as $O(1/n)$, yielding a vanishing value of irreversibility in the limit of large series. Roughly speaking, if we extend the relation between D and entropy production to the non-stationary realm, we would conclude that the process described in Equation (2.7) has a null lower bound for its entropy production $dS/dt \geq D_{kld}(\text{in}||\text{out}) = 0$, which is in good agreement with what is expected for a system which is in thermodynamic equilibrium.

The degree distributions for the VG have a power law decay k^{-2}, as reported in Figure 2.4. While we don't have a rigorous proof to support this, an heuristic derivation of this law can be outlined: the k_{out} of a node i (associated to x_i) chosen at random could be heuristically approximated as $P_{out}(k) \sim \#(k)q(k)$, where $q(k)$ defines the time window of the visibility basin (the average number of nodes that are 'visible' from i). As a rough approximation, $q(k)$ can therefore be related to the probability that the time series returns to x_i after an excursion where $x < x_i$, and this is of order $k^{-3/2}$ for unbiased random walks (the first return distribution of an unbiased random walker). On the other hand, node i won't necessarily have outgoing visibility will all and every node within the visibility basin, but just with a fraction of them. This fraction will depend on the fluctuations (roughness) of the time series within the basin. Roughness can be quantified in terms of the series standard deviation σ, which in unbiased random walks scale like $\sigma \sim t^{1/2}$. Accordingly, the percentage of k nodes visible within the basin of visibility should be of order $k^{1/2}/k = k^{-1/2}$. Summing up, $P_{out}(k) \sim k^{-3/2}k^{-1/2} \sim k^{-2}$, in good agreement with the results found in the panel (a) of Figure 2.4.

Finally, note that the finite size fluctuations decrease in VG at a slower rate than for HVG, scaling with series size as $O(n^{-1/3})$. This is perhaps due to the fact that degree distributions in the VG case are power laws instead of exponential ones, thus finite size effects in this case case decay slower than for HVGs.

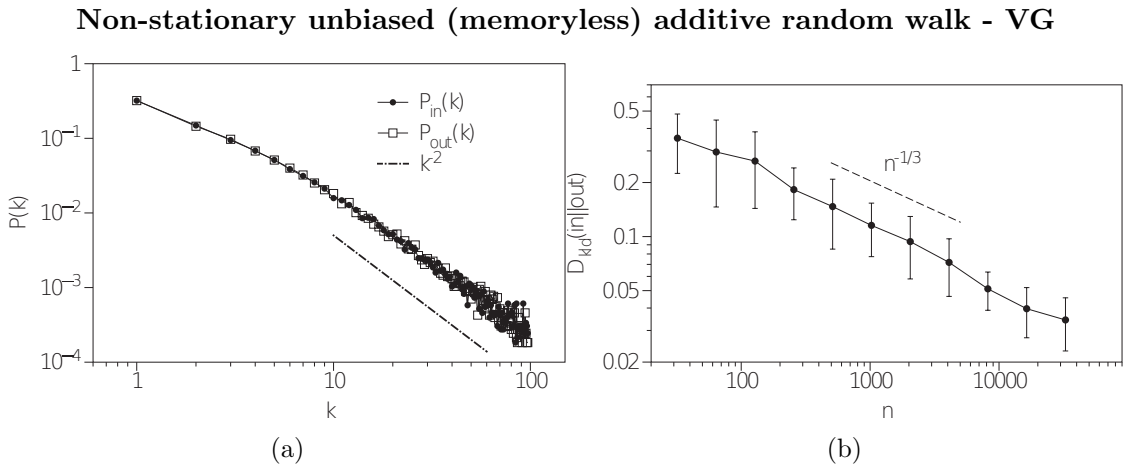


Figure 2.4: (a) Log-log plot of the in and out degree distributions of the natural visibility graph associated to an unbiased random walk of 2^{17} steps generated from $x(t+1) = x(t) + \xi$, where $\xi \sim U[-0.5, 0.5]$. Both distributions are identical up to finite-size effects fluctuations, suggesting that the underlying process is VG-reversible. The distributions follow a power law tail k^{-2}, something that can be heuristically justified according to scaling laws (see the text). (b) Log-log plot of the irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 10 realizations). This measure vanishes asymptotically as $n^{-1/3}$, suggesting that, although it is a non-stationary process, it is VG-reversible.

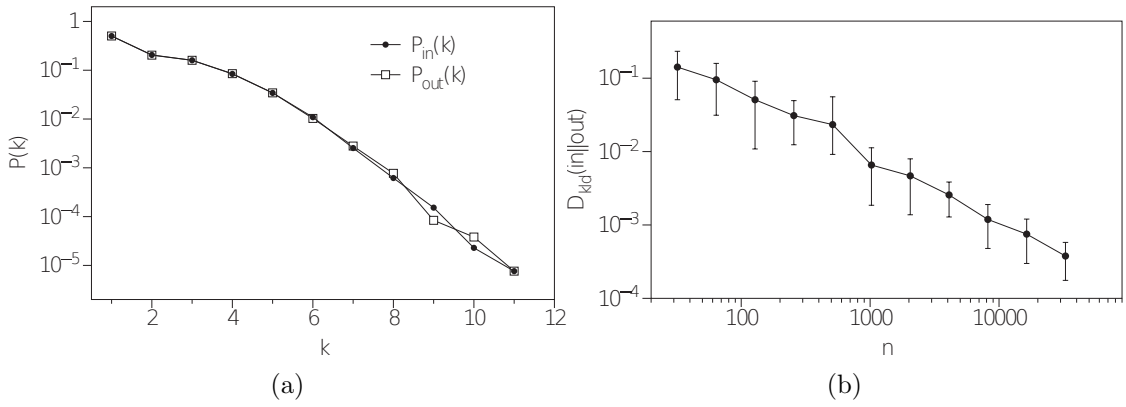


Figure 2.5: (a) Semi-log plot of the in and out degree distributions of the horizontal visibility graph associated to an unbiased random walk of 2^{17} steps generated from $x(t+1) = x(t) + \xi$, where $\xi \sim U[-0.5, 0.5]$. Both distributions are identical up to finite-size effects fluctuations, suggesting that the underlying process is HVG-reversible. (b) Log-log plot of the irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 10 realizations). This measure vanishes asymptotically as n^{-1}, certifying that, albeit being a non-stationary process, it is HVG-reversible.

2.3.3 Additive random walk with a drift

In this subsection we explore the effect of adding a positive drift to an additive random walk. For that purpose, we bias Equation (2.7) by defining its increments as having a small positive mean:

$$x(t+1) = x(t) + \xi,\ \langle \xi \rangle > 0$$

Note that this process is equivalent to superposing a linear trend, with positive slope $\langle \xi \rangle$ to the unbiased additive random walk described in Equation (2.7). Since the VG is invariant under addition of linear trends [3], the VG associated to an unbiased random walk and a random walk with a linear trend is the same, so again this process VG-reversible (of course, by symmetry something similar happens in the case of a negative drift $\langle \xi \rangle < 0$). Now, the HVG is not invariant under such transformation. Since the process is again VG and HVG stationary (choose $c = x_1 - x_{1+\tau}$ in Theorem 2.2.4), we should in principle be able to detect and quantify this additional source of irreversibility within the HVG setting. In Figure 2.6 we detail the numerical results for the HVG, for a concrete case where $\xi \sim [-0.4, 0.6],\ \langle \xi \rangle = 0.1$. The process is HVG-irreversible. As the method provides a finite positive irreversibility value that converges to $\lim_{n\to\infty} D \approx 7.5 \cdot 10^{-3}$, time asymmetry for this non-stationary process can be quantitatively distinguished from the unbiased case, for both finite and infinite size series. Extending again the analogy between irreversibility and entropy production to the non-stationary realm, the HVG method would provide in this case a tighter bound $dS/dt \geq D_{kld}(in||out) \approx 7.5 \cdot 10^{-3}$.

2.3.4 Non-Markovian additive random walk.

Finally, let us consider the following generalization of a random walk:

$$x_{t+1} = \begin{cases} x_t + \xi & \text{if} \quad p > r \\ x_{t-\tau} & \text{if} \quad p < r \end{cases} \tag{2.8}$$

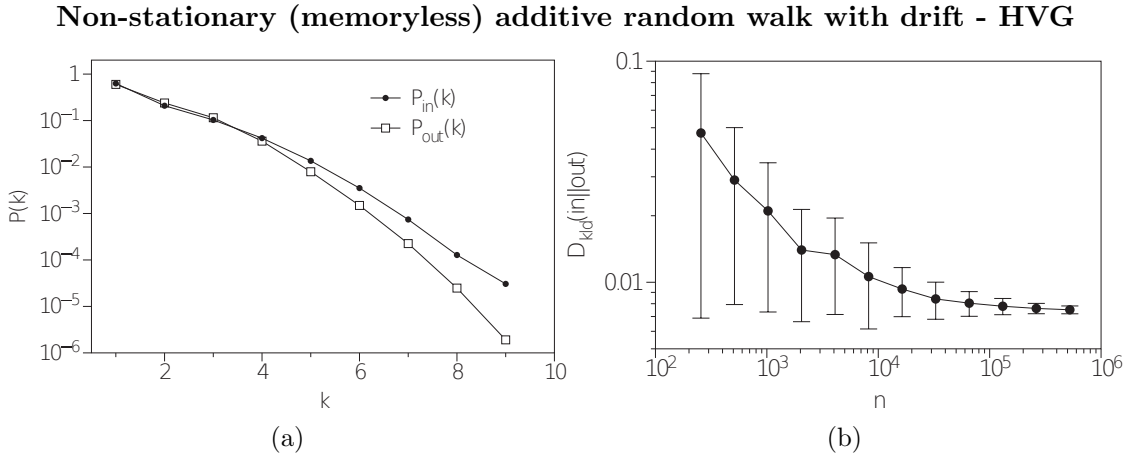
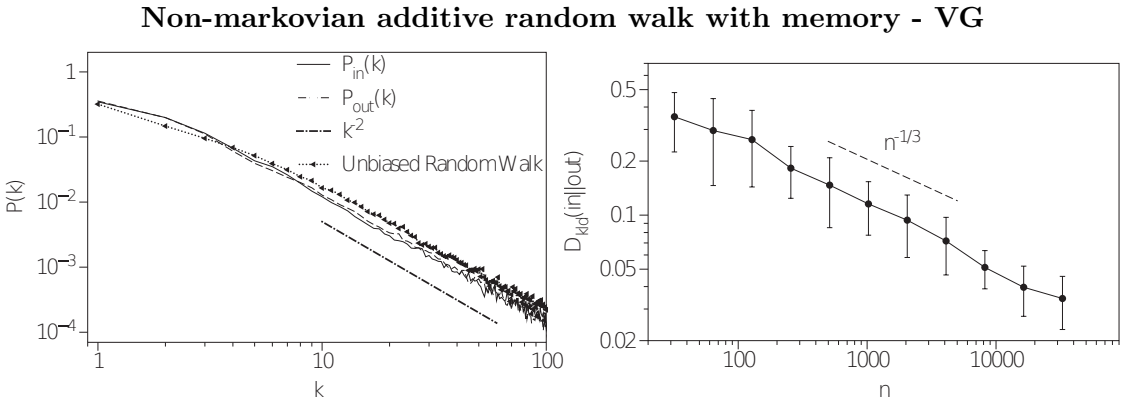


Figure 2.6: (a) Semi-log plot of the in and out degree distributions of the horizontal visibility graph associated to an unbiased random walk of 2^{17} steps generated from $x(t+1) = x(t) + \xi$, where $\xi \sim U[-0.4, 0.6]$, $\langle \xi \rangle = 0.1$. Distributions are different, suggesting that the process is HVG-irreversible. (b) Log-log plot of the irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 10 realizations, and error bars denote the standard deviation). This measure converges asymptotically to a finite value, certifying that the process is HVG-irreversible.

where for concreteness we set $\xi \sim U[-0.5, 0.5]$, $r \in [0, 1]$ is a fixed parameter that describes the reset rate, and $\tau \in \mathbb{N}$ is a fixed integer that describes the jump of the walker to previous states. This process can be used as a model for animal search in a 1d environment that includes memory of past locations [37]. It reduces to an unbiased Markovian random walk for $r = 0$, and is non-Markovian for $r > 0$ [37].

In order to investigate the capacity of these methods to capture irreversibility associated to off-equilibrium dynamics, we have computed the in and out distributions of both VG and HVG associated to the non-Markovian random walk described in Equation (2.8), for a specific time delay $\tau = 6$ and a resetting rate $r = 0.3$ (note that other values can be chosen as well). Results are shown in Figures 2.7 and 2.8. The system is clearly HVG-irreversible. As the unbiased case is HVG-reversible, the mechanism responsible of triggering the irreversible character is not the process non-stationarity, but the onset of memory effects that drive the system away from equilibrium, "producing entropy" at a rate $dS/dt \geq D_{kld}(in||out) \approx 8.8 \cdot 10^{-3}$.

On the other hand, the system appears to be VG-reversible again, failing to capture



(a) (b)

Figure 2.7: (a) Semi-log plot of the in and out degree distributions of the natural visibility graph associated to a biased random walk (see the text) of 2^{17} steps with delay $\tau = 6$ and reset rate $r = 0.3$. Both distributions are similar, suggesting that the onset of memory effects are not effectively captured by VG-reversibility, and although these are slightly different than for the baseline random walk, no major qualitative differences are observed. (b) Log-log plot of the irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 10 realizations). This measure vanishes asymptotically with series size as slowly as $n^{-1/3}$, so finite-size values can still be used for comparison with other models.

the source of irreversibility associated to the memory effects for asymptotic large sizes. However, note that the convergence speed of this process is rather slow (with $D \sim n^{-0.35}$ as reported in Figure 2.7), which permits us to compare finite-size irreversibility values. For instance, for $n = 10^3$, $D \sim O(10^{-1})$, compared with the much smaller analogous result for unbiased random walk $D \sim O(10^{-2})$. We can conclude that, although the VG fails asymptotically to detect irreversibility in this process, finite-size values can still be used in practice to compare the degree of reversibility with other processes.

2.4 Multiplicative random walks

We will now explore the properties of VG and HVG associated to certain random multiplicative processes. A multiplicative random walk (MRW) is a stochastic process $x(t)$ that follows the equation

$$x(t+1) = \xi \cdot x(t), \tag{2.9}$$

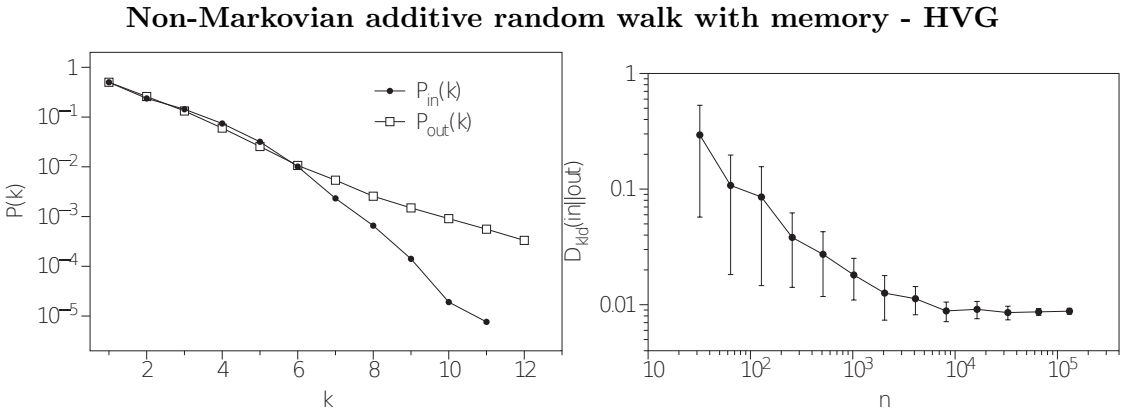


Figure 2.8: (a) Semi-log plot of the in and out degree distributions of the horizontal visibility graph associated to a biased random walk (see the text) of 2^{18} steps with delay $\tau = 6$ and reset rate $r = 0.3$. Both distributions are clearly different, suggesting that the onset of memory effects are effectively captured by the HVG. This seems to be a unique property of HVG (as VG fails to accurately capture this trait). (b) Irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 10 realizations). This measure converges with increasing series size to a finite, non-null value, certifying that the process is HVG-irreversible. As the unbiased (memoryless) random walk (Equation (2.7)) is in turn HVG-reversible, we conclude that the source of irreversibility captured in this process is only due to memory effects, as non-stationarities are filtered out.

where ξ is a random variable extracted from some distribution. If we identify $\mathcal{X} \equiv \log x$ and $\eta \equiv \log \xi$, this process is formally equivalent to an additive random walk in logarithmic space, as Equation (2.9) reduces to

$$\mathcal{X}(t+1) = \mathcal{X}(t) + \eta.$$

A standard assumption is that $\mathcal{X}$ approaches a log-normal distribution accordingly. However, we should be very cautious at this point, as the properties of an additive random walker, provided by the central limit theorem, are not directly applicable to the MRW in logarithmic space, due to non-ergodicity and to the relevant effect of extreme events [38], which preclude self-averaging and convergence to the asymptotic log-normal distribution. Also note that $\mathcal{X}$ only exists as the logarithm of x for $x > 0$, which in turn imposes restrictions on the support of ξ.

Case 1: Here we explore two simple versions of a MRW. In the first case we set $\log \xi$

Multiplicative random walk - Sample time series and degree distribution (1)

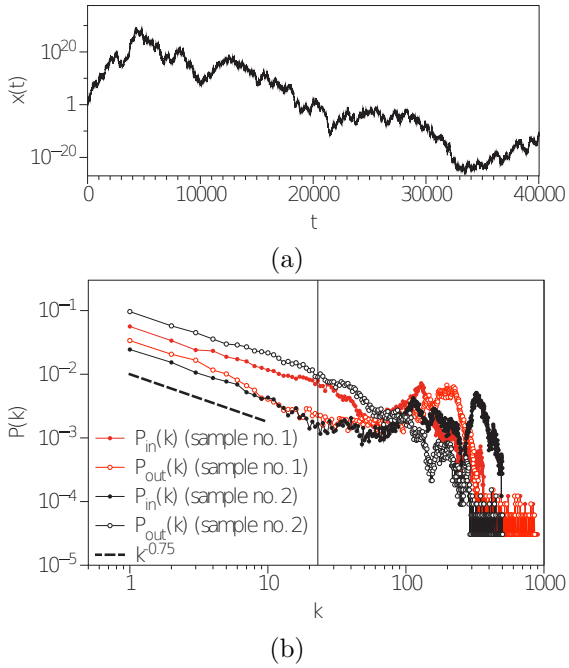


Figure 2.9: (a) Log-linear plot of a sample time series generated through the process $x(t+1) = \xi \cdot x(t)$, $\log \xi \sim U[-0.5, 0.5]$. This multiplicative process is additive in logarithmic space, hence in a log-linear plot, the time series looks similar to an additive random walk with no drift, as $\log \xi$ is uniformly distributed in a symmetric interval so $\langle \log \xi \rangle = 0$. Irreversibility measures on $x(t)$ are depicted in Figure 2.10. (b) Log-log plot of the *in* and *out* degree distributions for two different realizations of 2^{15} data of a multiplicative random walk $x(t+1) = \xi x(t)$, where $\log \xi \sim U[-0.5, 0.5]$. The curves have a power law decay with a fairly stable exponent $k^{-0.75}$, followed by wildly fluctuating tails. These are related to the presence of extreme events in the series, which are exponentially rare but exponentially large, and dominate the tails [38].

to be uniformly distributed in $[-0.5, 0.5]$, so $\xi \sim \exp(s - 0.5)$, $s \sim U[0, 1]$. With a positive initial condition for x, each realisation of this MRW is thus qualitatively similar to a realization of an additive unbiased uniform random walk in logarithmic space (see Figure 2.9 for an illustration). Notice that in the figure $x(t)$ wildly varies on amplitude, reaching values in the interval $[10^{-20}, 10^{20}]$ for $n = 4 \cdot 10^4$, however the sketch is, in log-linear scale, qualitatively similar to a realization of an additive unbiased random walk. At this point we need to introduce and prove an additional property of HVGs, starting with a definition.

Definition 2.4.1. A VG (HVG) is invariant under monotonic transformations if the VG (HVG) graphs associated to a time series $\{x_1, x_2, \ldots, x_n\}$ and $\{f(x_1), f(x_2), \ldots, f(x_n)\}$ (where $f(x)$ is an arbitrary monotonic function) are identical.

Proposition 2.4.2. VGs are in general not invariant under monotonic transformations.

Proof. The rationale is that the visibility link criterion is a convexity one, so any monotonic transformation that alters the convexity properties of the series will alter the resulting VG. We give here two counterexamples. Consider the time series $\mathcal{S}_1 = \{1, 2 \ldots, n\}$ and $f(x) = x^3$, such that $\mathcal{S}_2 = \{f(1), f(2), \ldots, f(n)\}$. The VG associated to $\mathcal{S}_1$ is a chain graph (1d lattice) whereas the VG associated to $\mathcal{S}_2$ is a complete graph. Consider now an additional series $\mathcal{S}_3 = \{\log f(1), \log f(2), \ldots, \log f(n)\} = 3\{\log 1, \log 2, \ldots, \log n\}$. Trivially, the VG associated to the concave series $\mathcal{S}_3$ is a chain graph , which is different from the VG of $\mathcal{S}_2$. □

Proposition 2.4.3. HVG are invariant under monotonic transformations.

Proof. The link criterion for HVGs is solely based on the specific *ordering* of the data, not on their values. Hence HVG shall be are invariant under order-preserving transformations. Monotonic functions are isotone mappings, hence order-preserving. Consider the time series $\mathcal{S}_1 = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{S}_2 = \{f(x_1), f(x_2), \ldots, f(x_n)\}$. The link criterion for HVG in $\mathcal{S}_1$ is: $x_k < \inf(x_i, x_j),\ \forall k : i < k < j$. But if this criterion is fulfilled, then we have $f(x_k) < \inf(f(x_i), f(x_j)),\ \forall k : i < k < j$, if f is monotonic. Therefore two connected nodes i and j in the HVG associated to $\mathcal{S}_1$ yield two connected nodes in the HVG associated to $\mathcal{S}_2$, which make both HVGs identical. □

In the light of the previous propositions, for $f(x) = \log x$, one finds that the HVG is equivalent to the HVG associated to a realisation of an unbiased additive random walk, i.e., the process is HVG-stationary and HVG-reversible. This is confirmed by the vanishing values of D in panel (b) of Figure 2.10 (ensemble averaged over 100 realisations). On

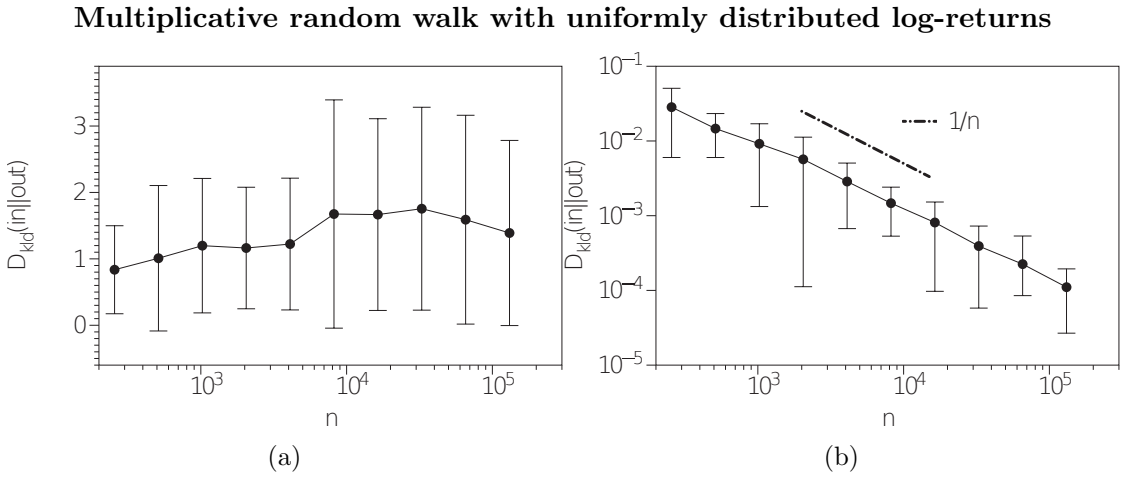


Figure 2.10: (a) Linear-log plot of the irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 100 realisations and error bars account for $\pm\sigma$) computed from the VG associated to a multiplicative random walk $x(t+1) = \xi \cdot x(t)$, $\log \xi \sim U[-0.5, 0.5]$. The measure converges to a finite value, so the process is VG-irreversible. (b) Log-log plot of the same measure computed from the HVG. The measure decays with series size n, so the process is HVG-reversible.

the other hand, by Section 2.4, VG is not in general invariant under log-transformations, so the VG does not reduce here to the one found in the additive case. In panel (a) of Figure 2.10 we plot the numerical results of D computed from the VG. First, D does not vanish with system's size, suggesting VG-irreversibility. Interestingly, values highly fluctuate within each ensemble average, as denoted by large standard deviations. This lack of self-averaging is related to the non-ergodic nature of the MRW: in panel (b) of Figure 2.10 we plot, in log-log scales, the *in* and *out* degree distributions of two different realisations of the process, for $n = 2^{15}$. The shape of the distributions all begin with a power law decay $k^{-0.75}$, followed by a tail that evidence large fluctuations. This is the part of the distributions that vary from realisation to realisation, and as it is ruled by extreme events (exponentially rare but exponentially different), deviations from the stationary distribution are large. This effect is well-known in multiplicative random processes [38], and precludes us to interchange average values with most probable ones (this is also the main reason why convergence of this process to the log-normal distribution is not straightforward). Because of that, distributions do not converge smoothly for large series to their asymptotic form, finding large standard deviations in the estimation of D.

Degree distributions of the VG for MRWs of varying lengths

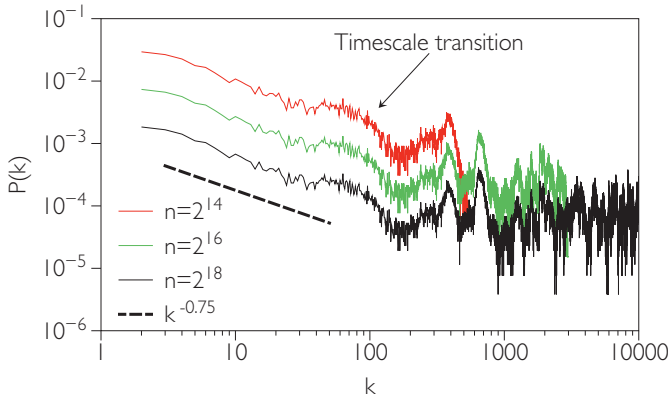


Figure 2.11: Log-log plot of the degree distribution of the VG associated to the first 2^n data (where, from up to bottom, $n = 14, 16, 18$) of a time series generated via the MRW with symmetric multiplicative noise $\xi \sim U[0.9, 1.1]$, which is qualitatively similar to an additive random walk with negative drift in logarithmic space. There is a power-law contribution at small degrees (associated to the fast time scale) and a steady wave-fluctuating part associated to the envelope whose extension increases with series size (see the text for details).

Case 2. In the second case, as an example of a MRW with symmetrical multiplicative noise, we set ξ to be uniformly distributed in $U[0.9, 1.1]$. With positive initial condition for x, each realisation of this MRW is qualitatively similar to an additive random walk with a small negative drift in logarithmic space (see Figure 2.12 for an illustration). The reason is that $\eta \sim \log s, s \sim U[0.9, 1.1]$, thus $\langle \eta \rangle < 0$, which implies that $x(t) \to 0$ for large values of t. This is a clear-cut case of an irreversible process. By virtue of proposition 2, the HVG should now be qualitatively similar to the additive random walk with drift case, hence HVG-irreversible. Indeed, we find that the process is finitely HVG-irreversible (panel (b) of Figure 2.13). On the other hand, the analysis based on VG is again not applicable (by Section 2.4) to the additive case. In order to explore the properties of VG, we need to advance yet another property:

Proposition 2.4.4. Let the time series $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ be such that there exist a convex function f such that $x_i = f(i)\ \forall x_i$. Then the VG associated to $\mathcal{S}$ is the complete graph $K(n)$. If instead of convex, f is concave, then VG is the chain graph (1d lattice).

Proof. Time series generated by convex functions generate complete visibility graphs. We can use a geometric proof. Consider two arbitrary data $x_i, x_j \in \mathcal{S}$ where without loss of generality $i < j$, and consider the segment that links x_i and x_j. As f is convex, the slope of this segment is always larger than the slope of any segment connecting x_i and x_k for $i < k < j$:

$$\frac{x_j - x_i}{j - i} > \frac{x_k - x_i}{k - i} \tag{2.10}$$

Equation (2.10) is equivalent to the visibility link criterion, so i and j are connected. Since this holds $\forall i, j \in [1, n]$, then the resulting visibility graph is the complete graph $K(n)$. The proof for the concave function follows analogously. □

Realistically, a time series extracted from the MRW with symmetrical noise cannot be represented as the graph of a convex function, and although there is a large-scale negative trend, there are also episodes of uphill fluctuations. However, as this process is equivalent to an additive random walk with a negative drift in logarithmic space, then, roughly speaking, the envelope of $x(t)$ can be approximated by $t^{-\delta}$, which is convex. This is a very crude approximation, and in reality we should expect that the degree distribution of VG is the result of two competing time scales. At slow time scales, there is a chunk of very highly connected nodes (which comes from the convexity part of the envelope). However, our time series is not strictly convex, and it (roughly) displays a composition of a convex function and a wave-like structure (associated to small scale uphill fluctuations). The random small scale fluctuations from the time series contribute with a random-walk like part, with distribution approximately equal to $k^{-0.75}$. These uphill fluctuations work as visibility barriers, hence we should not expect that the tail of the distribution follows $k^{-0.75}$, but instead a somewhat wavy distribution. As series size increases, a larger amount of data contribute to the envelope effect, hence we expect an increase of the slow time scale region and, by normalization, vertical shift in the distribution. We have run numerical simulations to explore this behaviour. We have generated a time series of 2^{18} data and computed the VG associated to time windows of the first $n = 2^{14}$; 2^{16} and 2^{18} data, and in Figure 2.11 we plot the degree distribution of

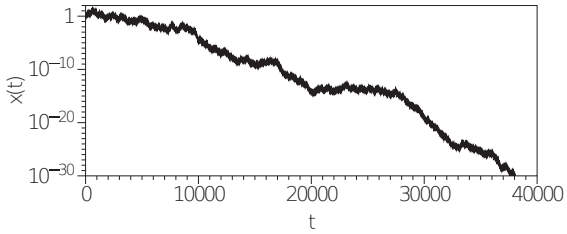
this VG for different series size. The two contributions as well as the transition between the two time scales can be clearly seen. Note that the shape associated to slow time scales will vary from realisation to realisation.

Regarding the directed VG, we expect that very few nodes have very large k_{out}, and these should correspond to early nodes associated to early data. On the other hand, a large percentage of the nodes will have large k_{in} (nodes associated to smaller and smaller values that receive links from past nodes). This percentage will increase as the series size increases, and will fluctuate in correspondence with the fluctuations of the time series, hence we expect large fluctuations from sample to sample. All these features are confirmed in panel (b) of Figure 2.12. On the other hand, the wild fluctuations at the level of the *in* degrees will generate finite irreversibility measures. Moreover, as the difference between *in* and *out* distributions are expected to increase with series size, the irreversibility measure should also be an increasing function of series size. This is confirmed in panel (a) of Figure 2.13, where the irreversibility measure increases with series size with no apparent bound, hence suggesting that the process is asymptotically infinitely VG-irreversible.

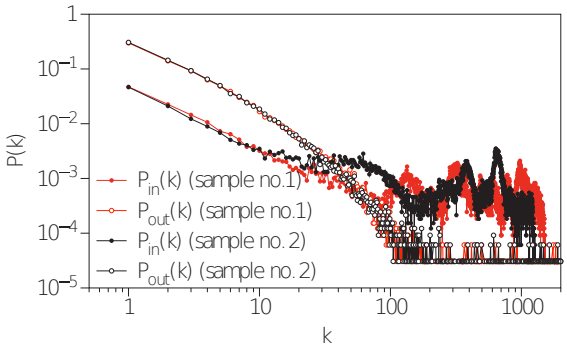
2.5 Discussion

In this chapter we have investigated, via analytical calculations and numerical simulations, the properties of visibility and horizontal visibility graphs associated to several non-stationary stochastic processes, as well as their ability to quantify several degrees of time irreversibility. We proved that unbiased additive random walks, while non-stationary, are both VG/HVG-stationary and VG/HVG-time reversible (remembering the fact that Brownian particles in thermodynamic equilibrium do not produce entropy on average). On the other hand, biased memoryless additive random walks are HVG-irreversible with finite irreversibility measures that quantify the degree of time asymmetry, while these are still VG-reversible, as VG is invariant under superposition of linear trends in the original data. Numerics suggest that HVG can capture for both finite and infinite series

Multiplicative random walk - Sample time series and degree distribution (2)



(a)



(b)

Figure 2.12: (a) Log-linear plot of a sample time series generated through the process $x(t+1) = \xi \cdot x(t)$, $\xi \sim U[0.9, 1.1]$, so in logarithmic space the process shares similarities with an additive random walk with a negative drift, as $\langle \log \xi \rangle < 0$. Irreversibility measures on $x(t)$ are depicted in Figure 2.13. (b) Log-log plot of the *in* and *out* degree distributions for two different realisations of 2^{15} data of a multiplicative random walk $x(t+1) = \xi x(t)$, where $\xi \sim U[0.9, 1.1]$.

size, the irreversible nature of non-Markovian additive random walks, whereas VG is only able to do so for finite series. For multiplicative random walks, the processes are HVG-reversible if the process is akin to an unbiased additive process in logarithmic space, and time irreversible if the process reduces to a biased additive process in logarithmic space. These latter results hold as HVG is invariant under monotonic transformations. Finally, the VG capture the time irreversible character of multiplicative random walks, yielding finite values in the unbiased case and asymptotically diverging quantities in the biased case.

While most of these are conclusions based on the asymptotic behaviour (i.e., in the limit of series size n), it should be noted that finite size time series always yield finite, non-null values of HVG and VG irreversibility. As the convergence speed for reversible processes

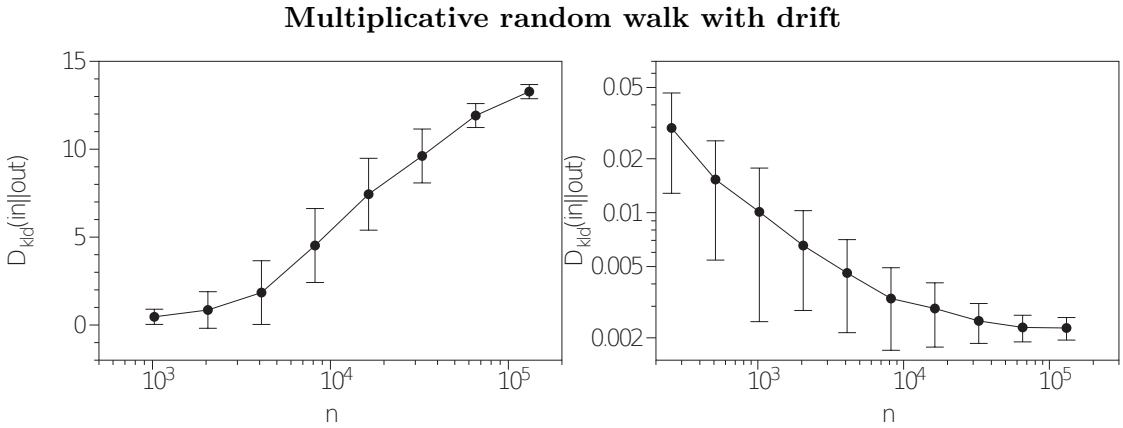


Figure 2.13: (a) Log-linear plot of the irreversibility measure $D_{kld}(in||out)$ as a function of the series size n (each dot is an average over 100 realisations and error bars account for $\pm\sigma$) computed from the VG associated to a multiplicative random walk $x(t+1) = \xi \cdot x(t),\ \xi \in U[0.9, 1.1]$, which induces an additive random walk with drift in logarithmic space. The measure diverges logarithmically with series size, hence the process is infinitely VG-irreversible. (b) Log-log plot of the same measure computed from the HVG. The measure converges to a finite, non-null value with series size n, hence the process is HVG-irreversible.

is rather slow ($O(1/n)$ for HVGs and $O(n^{-0.4})$ for VGs), these finite-size values can still be used in practice to compare the degree of statistical HVG/VG irreversibility for finite samples. This fact enables the use of VG/HVG methods in empirical (hence finite) datasets.

As a final remark, this analysis suggest that the horizontal visibility method seems to be better suited to capture irreversibility traits in additive non-stationary signals (I(1) processes), whereas the visibility method might be a better tool to quantify these signals which are better modelled by multiplicative processes. These results should be taken into account when assessing time irreversibility via visibility graphs, and will be useful particularly for the analysis of empirical non-stationary signals, such as financial, geophysical, or biological data.

Chapter 3

Visibility graph irreversibility in financial time series

3.1 Introduction

The quantitative analysis of financial time series [39] is a classical field in econometrics, which has received in the last few decades valuable input from statistical physics, nonlinear dynamics and complex systems communities (see [40–42]). In particular, the presence of long-range dependence, detected via multifractal measures has been used to quantify the level of development of a given market [43]. This approach was subsequently extended to address the problem of quantifying the degree of market inefficiency [44–48]. Here the adjective *efficient* refers to a market who is capable of integrating, at any given time, all available information in the price of an asset (the *weak-form* of the Efficient Market Hypothesis). In such a situation, assets should follow a martingale process in which each price change is unmodified by its predecessor, and the possibility of arbitrage would be impossible. Deviations from this ideal behaviour are thus characterising an inefficient system. In another approach, using the notion of forbidden patterns and permutation entropy, concepts arising in nonlinear dynamics, it was shown that one can

also evaluate the degree of inefficiency of a market [49]. These approaches are relevant from an applied perspective, as it has been shown that some links exist between the degree of inefficiency of a market and its predictability [50].

Another relevant property in the context of financial systems is that of time reversibility, which we explored in detail in Section 2.1. This concept of irreversibility has received some attention in the financial realm (for example see some initial investigations on this matter [51–55]), as has the concept of entropy [56]. However the amount of research in this area is limited, perhaps due to the fact that financial time series are usually non-stationary [39], and both the concept of time irreversibility as well as its relation to entropy production is not well defined in this case. As we discussed in Chapter 2, we can use the visibility graph to capture time series in such non-stationary systems, thus enabling us to study irreversibility in the financial realm. Accordingly, we first make use of the visibility algorithms to construct graph-theoretical representations of the stock prices of 35 companies from the New York Stock Exchange in the period 1998-2012. We then estimate time irreversibility in these representations through the Kullback-Leibler divergence of the *in* and *out* degree distributions. After checking that this measure is genuine and not correlated to volatility, we show that all the companies under study are irreversible, and their degree of irreversibility varies across companies and fluctuates over time. The variance across companies allows us to rank companies, and we use the collective time fluctuations to provide a classification of financial periods.

3.2 Methods

3.2.1 Irreversibility and entropy production

Recall the definition of the natural visibility graph (Definition 2.2.1 in Section 2.2), and the horizontal visibility graph (Definition 2.2.2 in Section 2.2). Recall also the definition of the Kullback-Leibler divergence (Definition 2.2.6 in Section 2.2). The concept of time series irreversibility, via KLD, thus reveals a deep link between information theory and statistical physics. In this chapter we aim at exploiting this link in the realm of finan-

cial systems. If one understands financial indicators evolving over time as the physical observables of the underlying (statistical-mechanical) system, then the irreversibility of these observables provides a lower bound on the system's dissipation. One can then consider the following questions: Is the financial system in "thermal equilibrium"? How do financial crisis and other major perturbations drive the financial system "away from equilibrium"? Which companies are evolving closer to equilibrium (and therefore are producing less entropy)? What is the relation between irreversibility and predictability, in the framework of financial series?

It is important to re-state at this point that financial time series are usually non-stationary. This is in principle a fundamental drawback. As a matter of fact, according to the original definition, non-stationary series are infinitely irreversible, so the quantification of how irreversible a non-stationary time series seems to be an ill-defined problem to begin with. Again, here we circumvent this problem by recalling our discussion in Chapter 2, where we used the visibility graph to quantify different degrees of irreversibility in both stationary and non-stationary processes.

3.3 Data and Results

We have analysed a dataset of financial stocks comprising stock evolution of 35 major American companies from the New York Stock Exchange (NYSE) and NASDAQ in the period 1998-2012, the majority of which belong to the Dow Jones Industrial Average (see Table 3 for a list). NYSE is the largest and most liquid cash equities exchange in the world by market capitalization. It therefore represents an appropriate set of observables to study the underlying evolution of the financial system. The series have very high resolution (intraday resolution of approximately one data per minute, in 1998-2012), yielding $\mathcal{O}(2 \cdot 10^6)$ data per company, which allows us to make a fine-grained and statistically robust analysis, and to break down data into different periods without losing statistical accuracy. We use the adjusted closing prices for concreteness. For illustration purposes, in Figure 3.1 we plot two sample series, representing the evolution of American

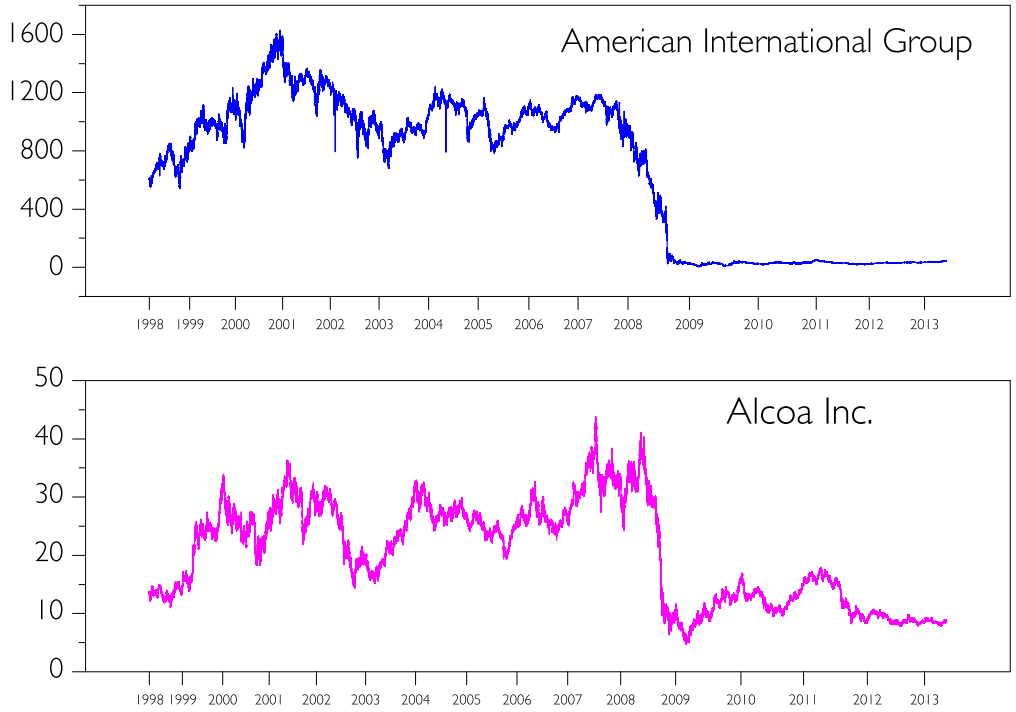


Figure 3.1: Stock price series $x(t)$ (adjusted closing price) as a function of time, for two different companies: Alcoa Inc. and Bank of America. Note that time is not strictly equispaced as there are some missing data for each company.

International Group and Alcoa Inc. (note that the series have been downsampled in this figure; each point is taken every 1000 time stamps). The dynamic range in this example differs greatly between these two companies, and this heterogeneity extends to the rest of the companies under consideration.

3.3.1 Basic measures of Irreversibility

In this section we explore and assess the irreversible character of financial data over the period 1998-2012. A priori, note that we can consider two different time series, namely the standard (non-stationary) price $x(t)$ and the log-returns $r(t) = \log(x(t)/x(t-1))$. This latter one is typically used in finance instead of $x(t)$, because it is believed that it is more stationary (a property of utmost importance for time series analysis [39]). Furthermore, because prices usually fluctuate by increasing or decreasing in terms of a percentage of the price, stochastic models of these price fluctuations acquire a more

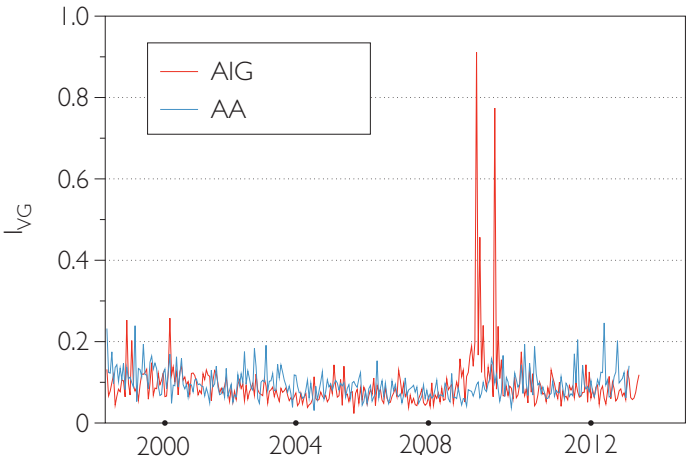


Figure 3.2: VG Irreversibility measure as a function of time (each dot represents I_{VG} associated to a non-overlapping time window of 5000 data points), for AIG (American International Group) and AA (Alcoa Inc.). The impact of the global financial crisis is only evident for AIG.

natural interpretation in logarithmic space. Historically it has been customary to use the rate of change $x(t+1)/x(t)$ as a random variable ξ. Thus $x(t+1) = \xi \cdot x(t)$ and this multiplicative process directly yields that $r(t)$ behaves as an i.i.d. random variable (white noise), which is stationary.

As we discussed in Chapter 2, visibility algorithms are well suited to study the irreversibility in non-stationary processes, hence it is not necessary to work with log-returns and we can directly use the original price series $x(t)$. As the underlying dynamics of $x(t)$ are expected to follow a multiplicative rather than an additive stochastic process, we shall focus on VG rather than HVG.

Our methodology is as follows: for each company c, we initially consider its price time series $x(t)$ all over the period 1998-2012. We then define a working time window of $n = 5000$ data points, and divide our original time series $\{x(t)\}$ of N data into a collection of N/n non-overlapping time series of n data each (we note that 5000 data points corresponds to approximately two weeks, which is a small enough period for us to make an analysis of how the time series evolves over the 12 years, but big enough to at least capture the dynamics). For each of these sub-series, we construct its associated VG,

and compute the irreversibility measure I_{VG} according to Equation (3.1) (results using the ℓ_1 norm are qualitatively similar), yielding a vector $I^c_{\mathrm{VG}}(\mathbf{w})$ for company c (where $\mathbf{w} = (w_1, w_2, \ldots, x_{N/n})$):

$$I_{\mathrm{HVG/VG}}(\mathbf{x}) := D_{\mathrm{kld}}(P_{\mathrm{in}}||P_{\mathrm{out}}). \tag{3.1}$$

We have thus computed $I^c_{\mathrm{VG}}(\mathbf{w})$ for all 35 companies in our dataset. All values obtained comply with the presence of VG irreversibility, as finite-size irreversibility values (for window size $n = 5000$) are generally higher than those found for reversible null models (white noise, additive random walks). This was expected, because as we discussed in Chapter 2, multiplicative models are known to display VG irreversibility, and this coincides with previous evidence [51–55]. Interestingly however, this quantity is fluctuating over time, and there are periods where the reversibility is comparable to additive random walks, thus we can state that in general stock prices are irreversible but periods of quasi-reversibility are not uncommon. Following the conceptual link between predictability and efficiency, one can infer that the market is more efficient when the stock price's reversibility approaches that found in reversible null models. In periods of financial stress, the irreversibility increases and the system is less efficient, and therefore somehow more predictable. For illustration purposes, the evolution of irreversibility for two companies, American International Group (AIG) and Alcoa Inc. (AA), is shown in Figure 3.2. One can immediately appreciate several differences among the two companies. For instance, whereas Alcoa Inc. (a metals technology corporation) seems to have a rather stable irreversibility over time, American International Group (a financial corporation) exhibits an abnormal irreversibility increase from year 2008, peaking in 2009. In principle one could ask whether there is a direct relation between the dynamic range of a certain series and its irreversibility. In the next section we rule out this possibility by observing a very small correlation between irreversibility and volatility.

The results for all companies are not shown but it is important to note that the amount of irreversibility is not a stable quantity, neither intra-company (i.e., for the same company,

over different periods), nor across companies. As pointed out above, intra-company irreversibility heterogeneity points out to the influence of exogenous factors, such as the impact of financially unstable periods. On the other hand, the fact that different companies have different irreversibility patterns is indicative that each observable of the financial system evolves over time in a different fashion, and thus this property can be used to rank companies accordingly. In the next sections we investigate these aspects.

3.3.2 Ranking Companies

In order to quantify the net amount of irreversibility of a certain company, we introduce Score[c], the score of a company c as the average of the annualized irreversibility value

$$\text{Score}[c] = \frac{1}{15} \sum_{\text{year}=1998}^{2012} I^{c}_{\text{VG}}(\text{year}). \tag{3.2}$$

This quantity averages the degree of irreversibility of a given company over large periods of time. According to the analogy between reversibility and entropy production, the larger the Score is, the more "away from equilibrium" the signal generated by c is, thus producing larger amounts of entropy. This might be relevant from a financial perspective, as the larger the Score of a company, the less efficient it is and thus more interesting from an investment viewpoint. The companies can be ranked accordingly and the results are depicted in Table 3. No obvious interpretation can be stated at this point, as one finds multinationals operating in different sectors (insurance, industry, health) in the top six ranks.

We now compare the new defined metric with standard financial metrics; we make use of the annualized volatility, which is commonly used to capture the dynamic range of financial data. We will define volatility as the standard deviation of the price log-returns over a year. In Figure 3.3 we plot, for each company, the averaged annualized volatility (defined as the average of annualized volatilities over 1998-2012) against its Score. If both measures were correlated, we would expect that a smooth curve emerges in the scatter plot. The scatter, however, is large. The red line describes the best fitting of the data

Acronym	**Name**	**Comment**	**Score Rank**
GM	General Motors Company	NYSE	1
AIG	American International Group, Inc.	NYSE	2
TRV	The Travelers Companies, Inc.	NYSE	3
AA	Alcoa Inc.	NYSE	4
UNH	UnitedHealth Group Incorporated	NYSE	5
MO	Altria Group Inc.	NYSE	6
HPQ	Hewlett-Packard Company	NYSE	7
HD	The Home Depot, Inc.	NYSE	8
CVX	Chevron Corporation	NYSE	9
MRK	Merck & Co. Inc.	NYSE	10
CAT	Caterpillar Inc.	NYSE	11
MCD	McDonald's Corp	NYSE	12
KO	The Coca-Cola Company	NYSE	13
C	Citigroup Inc.	NYSE	14
DD	E.I. du Pont de Nemours and Company	NYSE	15
IBM	International Business Machines Corporation	NYSE	16
VZ	Verizon Communications Inc.	NYSE	17
BA	The Boeing Company	NYSE	18
PG	The Procter & Gamble Company	NYSE	19
HON	Honeywell International Inc.	NYSE	20
JPM	JPMorgan Chase	NYSE	21
GE	General Electric Company	NYSE	22
UTX	United Technologies Corporation	NYSE	23
CSCO	Cisco Systems, Inc.	NYSE	24
AXP	American Express Company	NYSE	25
XOM	Exxon Mobil Corporation	NYSE	26
T	AT&T, Inc.	NYSE	27
BAC	Bank of America Corporation	NYSE	28
PFE	Pfizer Inc.	NYSE	29
INTC	Intel Corporation	NasdaqGS	30
DIS	The Walt Disney Company	NYSE	31
MMM	3M Company	NYSE	32
WMT	WMT&Wal-Mart Stores Inc.	NYSE	33
MSFT	Microsoft Corporation	NasdaqGS	34
JNJ	Johnson & Johnson	NYSE	35

Table 3: List of companies and associated irreversibility Score Rank (see the text).

to a linear relation, with a poor Pearson's linear correlation $r^2 = 0.06$. We conclude that volatility and irreversibility are not correlated, making the latter a genuine and complementary metric.

To further assess the possibility that some companies may have suffered from large irreversibility only at sporadic occasions (which would yield a high irreversibility score

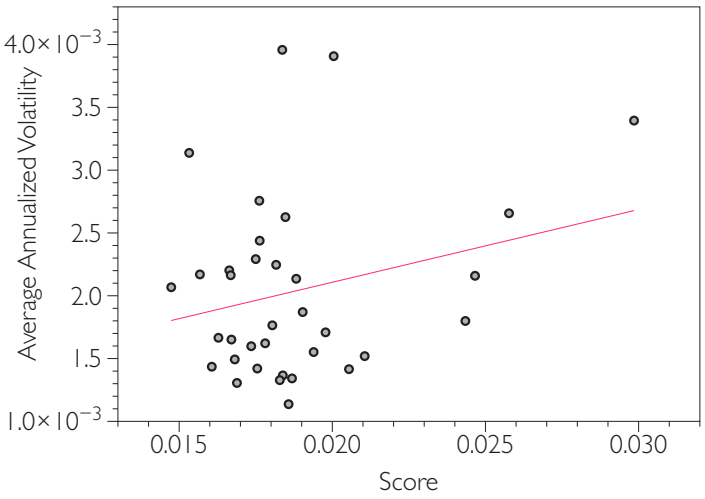


Figure 3.3: Scatter plot of the Score against the average annualized volatility of each of the 35 companies. The figure is highly scattered (the solid red line provides the best fitting to a linear dependence between both quantities, with Pearson's $r^2 = 0.06$). This indicates that volatility and irreversibility Score are not correlated, suggesting that the latter is a genuinely new and complementary measure that provides different information about the evolution and performance of a given company.

even if the company were following a quasi-reversible evolution in most of the period), we also compute the irreversibility variance $\sigma^2(D_{\text{kld}})$ of a given company, with

$$\sigma^2(D_{\text{kld}}) = \langle I_{\text{VG}}^2 \rangle_{\text{years}} - \langle I_{\text{VG}} \rangle_{\text{years}}^2. \tag{3.3}$$

A company with low irreversibility variance corresponds to one whose evolution is relatively independent of the particular strength of exogenous variables; its response against external perturbations is relatively independent of the strength of the perturbation, and only depends on endogenous properties of the company. This is typically the case for a system composed by quasi-uncoupled variables evolving over time and subject to random perturbations with a well-defined perturbation mean (for instance, Gaussian perturbations). On the other hand, if a company has a large irreversibility variance, its response varies largely with the type of perturbation. This is usually the case either for uncoupled variables subject to random perturbations with fat tail distributions (for example, power law distributed perturbation strength), or for highly coupled variables. For instance, AA (Alcoa Inc.) has large a irreversibility Score, however reasonably low irreversibility vari-

ance according to Figure 3.2, hence the dynamics underlying the evolution of the price of this company have not dramatically changed over time (i.e., external perturbations approximately have the same impact). On the other hand, AIG (American International Group) shows a relatively large irreversibility score and variance. Hence AIG, over the period 1998-2012, was affected differently by different financial perturbations. In particular, the perturbation that originated in 2008 has a qualitatively larger effect on the company's internal dynamics than other perturbations; this can also be seen in Figure 3.2.

In general, the irreversibility Score will be a faithful static measure of a company's irreversibility as long as we have relatively small variance. We have compared these measures for all companies in Figure 3.4. We find a bulk of companies for which the Score ranges between 0.015 and 0.020, where the variance is relatively constant. This means that the Score alone is a sufficient indicator of irreversibility, at least for these companies. Interestingly, we find that the top six multinationals in the Score ranking, also have large variance. These are companies which have been dramatically affected by major external perturbations at certain specific times, perhaps acting as global sensors of the financial system's stability state. In the next section, we further explore this possibility, and investigate, in an unsupervised way, if the evolution of irreversibility features across companies over time reflects the stability of the whole financial system. This allows us to classify and cluster periods of time according to their level of systemic reversibility.

3.3.3 Assessing periods of financial reversibility

In order to be able to quantitatively compare the performance of the whole system amongst different periods of time, we consider the quantity $I^c_{\mathrm{VG}}(\mathrm{year})$, which is the irreversibility in company c in the respective year, and create the vector $I^{\mathbf{c}}_{\mathrm{VG}}(\mathrm{year})$ where $\mathbf{c} = (c_1, \ldots, c_{35})$ is the vector of companies. Our entire database is then coarse-grained into a set of 15 (the number of years), 35-dimensional observations (the companies)

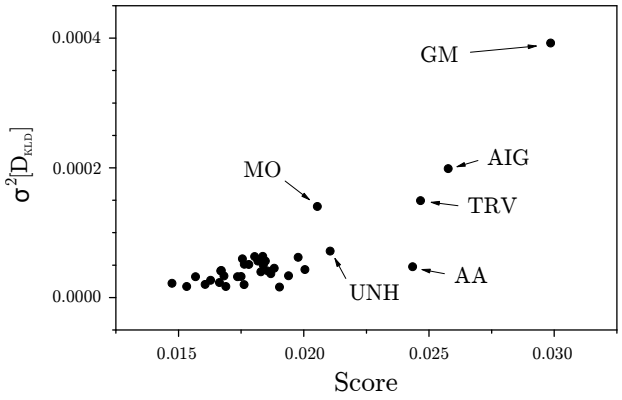
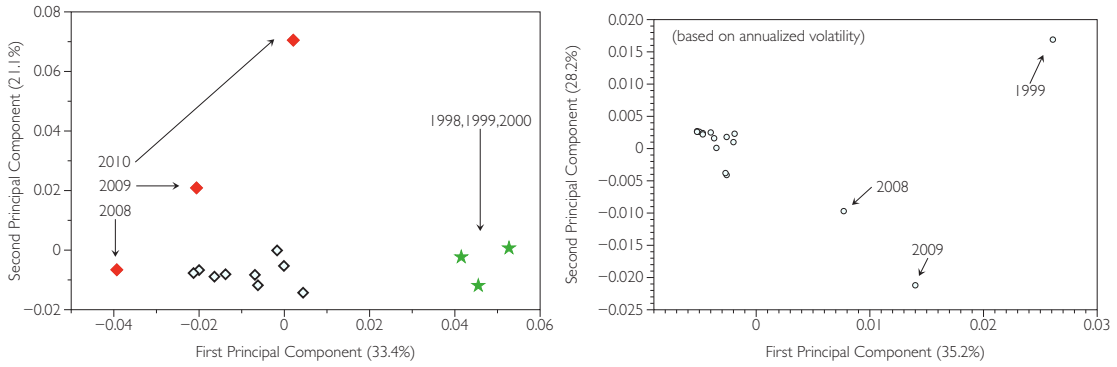


Figure 3.4: Scatter plot of the irreversibility variance against the irreversibility score for each company. The top six companies from Table 3 are highlighted.



(a) (b)

Figure 3.5: (a) Projection of financial periods in the PCA space of $\{I^{\mathbf{c}}_{\mathrm{VG}}(\mathrm{year})\}^{2012}_{\mathrm{year}=1998}$. The first two principal components account for about 55% of the system's variability. In the two dimensional space spanned by these components, we can already find three clusters, which account for somewhat stable years, dot-com bubble and the maximum of the global financial crisis. (b) Similar analysis but using annualized volatilities. While results show some qualitative agreement, the analysis based on irreversibility provides a clearer picture.

spanned by $\{I^{\mathbf{c}}_{\mathrm{VG}}(\mathrm{year})\}^{2012}_{\mathrm{year}=1998}$. In order to find patterns arising among different periods, we make use of two standard techniques in data mining: principal component analysis (PCA) and hierarchical clustering.

Principal Component Analysis (PCA) [57] is a common statistical procedure to perform dimensionality reduction on data. It uses an orthogonal transformation to project our set of observations (originally described in $\mathbb{R}^{35}$, where each direction is possibly correlated

among observations, as similar companies might have correlated irreversibility evolutions) into a lower dimensional subspace spanned by the principal components, obtained from the eigenvectors of the dataset covariance matrix. These particular directions are such that (i) they are orthogonal, (ii) the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. Thus projecting each observation O_i (originally $O_i \in \mathbb{R}^{35}$) into a smaller space spanned by the first m principal components hugely reduces the dimensionality of the observations, while keeping the relevant information of the data. This projection is the one that minimises the mean squared distance between the data points and their projections.

In the left panel of Figure 3.5 we show the projection of $\{I^{\mathrm{c}}_{\mathrm{VG}}(\mathrm{year})\}^{2012}_{\mathrm{year}=1998}$ in the PCA space spanned by the first two principal components (accounting for about 55% of the data variability). Interestingly, observations (i.e., years) automatically seem to cluster into three separated groups. The first one includes the observations for years 1998, 1999 and 2000, a period that can be identified with the dot-com bubble. The second group includes the years 2008-2010, which is in turn well known to represent the period of largest financial stress resulting from the global financial crisis. The third group amalgamates the rest of the years, but it is difficult to find finer structures within that one with this representation. Note that these results are on good agreement with recent, alternative metrics that make use of multiplex visibility graph mutual information [58].

For the sake of comparison, in the right panel of Figure 3.5 we have performed a similar analysis, but using the annualised volatilities of each company as the features of the vectors characterizing each year, instead of the average irreversibilities. First, note that while irreversibility and volatility are not correlated (see Figure 3.3), we obtain qualitatively equivalent results in PCA space. This suggests that both measures provide complementary information. However, by only using volatilities we seem to be losing detail as both financially unstable periods are under-represented in this latter case.

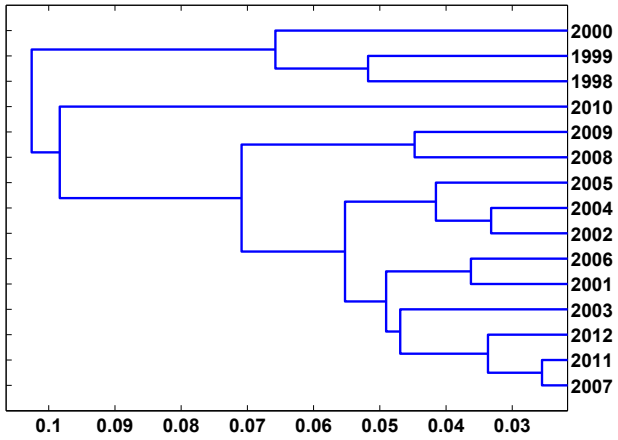


Figure 3.6: Hierarchical clustering obtained from $I^{\mathbf{c}}_{\mathrm{VG}}(\mathrm{year})$. Years 1998, 1999 and 2000 group together at the top (analogous to the cluster seen in the left panel of Figure 3.5, which can be interpreted as the dot-com bubble). Years 2008 and 2009 also group together.

Since the two dimensional space spanned by these projections only account for 55% of the data variability, it is not totally straightforward that this particular two dimensional projection is *faithful*, that is to say, we need additional evidence to confirm that any emerging separation or clustering is not spurious. In order to further explore in more detail the fine grained relation between years, we then make use of hierarchical clustering. This is a method of cluster analysis which seeks to build a hierarchy of clusters, where the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented in a dendrogram. To build this, we first compute a distance matrix d among observations, where d_{ij} is in this case the L_2 (Euclidean) distance between observations i and j in the original high-dimensional space $\mathbb{R}^{35}$:

$$d_{ij} = \sqrt{\sum_{k=1}^{35} \Big(I_i(c_k) - I_j(c_k) \Big)^2}$$

A complete linkage criterion is considered to generate an agglomerative hierarchical cluster tree from the distance matrix d. In Figure 3.6 we plot the dendrogram generated via hierarchical clustering based on the $I^{\mathbf{c}}_{\mathrm{VG}}(\mathrm{year})$. This provides more visibly refined information compared to projection in PCA space. When we move left in the dendrogram,

we are coarse-graining details. For instance, we can see that years 1998, 1999 and 2000 group together (analogous to the cluster seen in the left panel of Figure 3.5, which can be interpreted as the dot-com bubble). At the same level of granularity, 2008 and 2009 group together (which falls in the period of the global financial crisis). As we decrease the granularity, 2010 pairs up with 2008-2009, leaving a bulk of other years which can be identified with more stable and efficient financial periods. Interestingly, if we look at smaller scales, we also find that years 2011 and 2007 group together, suggesting that the onset and exit of the global financial crisis leaves a similar fingerprint on the irreversibility metric. In conclusion, these results are in good agreement with those found using PCA, and suggest that the evolution of the system's financial stability can be extracted from the collective evolution of companies irreversibility features.

3.4 Discussion

In this chapter we have extended and studied the concept of time irreversibility to the context of financial time series. While this is a statistical concept which can be linked with entropy production in (non-equilibrium) stationary states, it can be extended to the realm of *non-stationary* time series by using the visibility algorithms. These transform time series into graph-theoretical representations and allow for a direct quantification of time series irreversibility even if the associated dynamical process is non-stationary.

We found that the stock prices of the companies in our dataset are time irreversible, in the sense that finite-size irreversibility values are higher than those found for reversible null models. This is yet more evidence that exposes the inefficiency of financial systems, supporting the violation of the classical efficient market hypothesis. It is however important to note that different companies have distinct time evolving irreversibility patterns, and some display periods of quasi-reversibility, which implies that (i) some companies are *more* irreversible than others, and (ii) the degree of reversibility of each company varies over time. According to (i), one can rank companies. As there is a conceptual link between predictability and efficiency, one can argue that reversible time series are

less predictable than irreversible ones. In this sense, the ranking of companies based on stock price irreversibility could provide relevant information for traders and optimal portfolio designs, and we have shown that this information differs to that gained from volatility measures. According to (ii), one can also rank periods of financial stability. Concretely, we found that periods of financial turmoil, such as the dot-com bubble or the global financial crisis, can be easily identified and distinguished from periods of financial stability if we use the irreversibility values of each company as the features to feed clustering algorithms.

We conclude that the concept of time irreversibility, adequately adapted to a financial time series scenario via visibility algorithms, reveals complementary and valuable information on the evolution and the structure of stock prices. Further research is needed to explore the relation between irreversibility, predictability, or more generally, the evolution and dynamics of stock prices.

Chapter 4

Peak/pit asymmetries in visibility graph space

4.1 Introduction

A major challenge in studying temporally unfolding natural systems is making sense of data that are noisy, reflect processes at local and global scales, and are likely non-stationary. This is exemplified in diverse domains in the natural sciences where theoretically driven research aims to describe how system organisation and interaction dynamics relate to produce time series features [39, 59]. Researchers therefore require time-series descriptors that are: highly informative while still offering a good compression, relatively robust against instrument induced fluctuations, and (particularly in neurobiological domains) invariant to individual differences that exist between samples.

Particularly important for research across disciplines including neurobiology, finance and earth sciences is that such descriptors differentiate the contribution of oscillatory processes from that of point-like impulses that are inherently non-oscillatory. This is apparent when considering spectral power measures, where both oscillatory processes and regularly spaced point-like events produce similar signatures. Additionally, the structure

of point-like events could itself reflect different processes that determine local minima and maxima in a time series. This translates into a need to quantitatively measure the anisotropy of a fluctuating signal in a manner that captures to the difference in the structure of peaks (local maxima) and pits (local minima). Departing from analytic approaches based on consideration of an entire time series (captured by variants of spectral analysis, entropy, or non-linear signatures) one that contrasts structure within local peaks and pits is certainly needed, and we present such an approach here.

To further motivate the aforementioned theoretical challenge, we shall discuss the issue in some concrete disciplines. In neurobiology, and more specifically within research examining the brain's spontaneous mode of operation during wakeful rest ("resting state processes"), it has long been known that Blood-oxygen-level-dependent (BOLD) time series in the human brain show strong power at low frequencies (< 0.1Hz) and that such frequencies underlie resting-state connectivity [60]. This has produced an entrenched view of natural brain dynamics: these dynamics are taken to reflect slow, synchronous oscillations at low frequencies. However, more recent research is shedding light on the problems with such descriptions. For instance, [61] identified strong and highly infrequent spontaneously occurring spike-like events within recordings of resting state brain activity. They showed that removing these events from the time series reduced low-frequency power by as much as 60% in certain brain areas. This removal also reduced correlations within brain networks by as much as 50%. Other work suggests that such infrequent spontaneous events carry much weight in explaining important phenomena related to brain function. In particular, Taglizaucchi et al. [62] showed that well-studied connectivity networks in the human brain can be accurately reconstructed even keeping *only* 1% of the data in each time series (for related work, see [63]).

In parallel with the shift towards consideration of the role of rare, relatively extreme events, other work began examining in more detail the specific features of both local maxima and local minima in neurobiological time series. Many neurobiological time series, including those measured by fMRI scanners, do *not* have a natural minima where

the measured signal is zero. From the neurobiological perspective, the need to understand minima and maxima in such cases emerges naturally within any model in which oscillatory patterns or power spectra are not the only source of information on brain activity.

To illustrate, in the auditory system, activity peaks are known to track physical features of the input (e.g., frequency or amplitude), so that external stimuli produce well defined steady-state responses (peaks in activity spectra) that track time-varying features of auditory inputs (e.g., [64]). However, within the same auditory system, activity pits may be further impacted by other factors, including "dampening" induced by visual processing, the level of overall attention paid to the auditory stimuli, or the degree of engagement in memory maintenance. This is a simple example of how local maxima and local minima may provide information about different processes. Additionally, studies of resting state neuroelectric responses in the human brain have distinguished between modulations of oscillatory peaks from modulation of oscillatory troughs. That line of work has documented a difference between the variance of peaks and pits in resting-state time, (Amplitude Fluctuation Asymmetry [65]), linking these to activity in visual areas, and suggesting this asymmetry derives from unbalanced forward vs. backward propagating currents within dendrites. Studies of functional Magnetic Resonance Imaging (fMRI) have also examined asymmetries in resting-state activity. These studies first identified all local minima and maxima in each time series and then contrasted the respective variances of the sets of local maxima and local minima. This variance asymmetry within spontaneous brain activity distinguishes adults from children [66] and differentiates between wakefulness and sleep stages [67]. As noted by Mazehari and Jensen [65], "the amplitude fluctuations of oscillatory activity are conventionally viewed as being symmetric around zero", but as we summarized above, emerging findings show that these conventions require revision, and importantly, that new and precise analytic tools are needed to quantify features of asymmetry within local minima and maxima. As we later show, amplitude asymmetry might be less than ideal in identifying such dynamics.

The need to capture and quantify possible asymmetries between the local maxima and minima of time series is not just inherent in neurobiological processes. Let us consider the time evolution of some financial index $x(t)$ (e.g., S&P500) which represents the aggregate, collective evolution of the interaction of a number of financial assets over time. In quantitative finance, it is a well-known empirical fact that qualitatively different dynamics operate microscopically when $x(t)$ is on average increasing with respect to when this index is in a sustained decrease (which is equivalent to say that $-x(t)$ grows). In the former case, market is usually stable, pairwise correlation between the constituting assets is generally low (the system is said to be close to equilibrium [6]) and risk perception is low, leading to a time series with low volatility. Conversely, a situation where $x(t)$ decreases is indicative of a market under stress, where the correlation of the constituent assets grows due to common factors. As a result, any small and uncontrollable fluctuation can easily propagate throughout the system, hence the dynamics display larger volatility (larger uncertainty). The role that these two different market dynamics is playing can be therefore examined by analysing series statistics under index inversion $x(t) \to -x(t)$, something that we will show is tightly related to asymmetries between peak and pit statistics.

Keeping in mind the aforementioned desiderata and as well as findings highlighting the importance of discrete events, our aim here was to develop a general method for the efficient quantification of peak/pit asymmetry targeted at understanding the role of local, non-oscillatory processes. After a thorough validation of such a methodology, our aim was then to apply it in a wide range of settings, including neurobiological, financial, climate time series and beyond. We capitalized on a recent general approach to the description of time series that provides information about both local and global temporal features, without assuming neither stationarity nor oscillations at any temporal scale.

This approach originates on the notion of a visibility graph, which we introduced and formally defined in Chapter 2, and is a running theme in this thesis. Importantly, a fun-

damental property is that the conversion from time series to visibility graphs is invariant under several transformations that map onto to common nuisance (e.g., instrument) effects which are typical in neurobiological time series and beyond. These include linear trends, amplitude modulation on longer scales, or variations (slowing-down, speeding-up) in the rate of the process in question. Consequently, these invariances result in more efficient combination of information across measurements after transforming a time series into a visibility graph. Visibility graphs provide informative features both at the local level of single vertices and at the global level, where distributions of vertex features are described. Such global features reflect, for instance, the self-similarity of fractal time series [25], estimation of entropy production due to time irreversibility [18], discrimination between noise and chaos [33, 68], etc. Interestingly, previous works report that visibility graph features capture both linear and nonlinear information of the dynamical process and thus extend above and beyond standard power spectrum-like measures which only capture linear correlations.

Additionally, because visibility graphs can provide insights into local, non-oscillatory processes [69], they go beyond the information provided by "global" measures that summarize time series in a single parameter [59] (series entropy, fractal dimensions, power spectrum or even methods that are sensitive to similar (repetitive) motifs on multiple time scales (e.g., multiscale entropy methods)).

Equipped with the notion of visibility graph as a starting point, we will explore a systematic extension to that method, which is designed to satisfy the desiderata outlined above and capture different dynamics that may determine peaks and pits in a given signal. As indicated above, we are interested in a measure of asymmetry that can be applied to time series that do not have a natural minimum, which can be combined across measurements, and which is relatively robust to noise.

The rest of this chapter proceeds as follows: in Section 4.2 we introduce the theoretical formalism, along with a theoretic analysis and validation for synthetic processes. For simple (stochastic and deterministic) processes we show that this methodology can cap-

ture peak/pit asymmetries with similar performance to that existing indicators, however for more complex processes involving combination of different dynamics and scales we show that this methodology outperforms existing approaches. We confirm such findings with extensive numerical simulations and rigorous results on concrete, canonical complex dynamics. In Section 4.4 we first show that this method offers novel descriptions for spontaneous brain activity in humans and differentiates between states of consciousness. For financial time series, it captures important features of financial regimes linked to major events in stock markets. We finally explore the application of this framework for extensive climatic data.

4.2 Methods

4.2.1 Graph-theoretical framework for peak/pit asymmetry quantification

Let us consider a real-valued time series of size N, $\mathcal{X} = \{x(t)\}_{t=1}^{N}$. The traditional step to get access to peak and pit statistics is to define two ordered sets, namely $peak = \{x(t)|x(t+1) < x(t), x(t-1) < x(t)\}$ and $pit = \{x(t)|x(t+1) > x(t), x(t-1) > x(t)\}$. The hypothesis is that the statistics of these sets, and its difference, carry relevant information on local fluctuations of $\mathcal{X}$ and can be used as a feature for making diagnostics. Mathematically, differences in the statistics of *peak* and *pit* can be related in principle to two scenarios, namely: ($\mathcal{S}_1$) different marginal distributions, and ($\mathcal{S}_2$) different correlations (i.e., *peak* and *pit* might have similar marginals but different temporal correlations). Additionally, ($\mathcal{S}_3$) characteristic cross-correlations between *peak* and *pit* can also be informative (e.g., *peak* and *pit* might have similar marginal distributions but say, fluctuate in an anti-correlated way with one another). In Figure 4.1 we plot a example time series showing (very crudely) the three scenarios. Among the plethora of possible descriptors, one should be able to identify and separate which of the scenarios ($\mathcal{S}_{1-3}$) the measures are addressing. Asymmetries between peak and pit statistics have only been addressed relatively recently, and existing statistical tests e.g., Amplitude Fluctuation Asymme-

try (AFA [65]) or Amplitude Variance Asymmetry (Davis2014 [66]) essentially considers scenario ($\mathcal{S}_1$) by comparing the variances of these marginals (through the logarithm of the Variance Ratio VR $= \sigma^2(\text{peak})(t)/\sigma^2(\text{pit}(t))$ in the case of AVA, where $\sigma^2(X)$ is the variance of the random variable X). As these measures only considers the one-point marginals of each set, they cannot give us particular insights on scenarios ($\mathcal{S}_2$) or ($\mathcal{S}_3$), and therefore disregard these aspects. However it is a computationally simple statistic, and probably because of its simplicity the quantity $|\log(\text{VR})|$ is currently used to assess the similarity between peak and pit statistics as mentioned above when describing prior applications to neurobiological time series. In addition, the degree to which AVA or AFA in any given time series is driven by extreme points [70] is unclear and necessitates manual examination. In practice, it appears there may be a relation between the AVA of brain time series and the skewness of the empirical distribution [70], though analytically AVA and skewness are independent as the same distribution can produce time series with markedly different AVA values depending on specific sampling. Incidentally, note that the quantity $|\log(\text{VR})|$ fulfils the axioms of a metric only in the case where $\log(\text{VR})$ is positive. This can be easily proved from the triangle inequality: take X, Y, Z. For $\sigma^2(X) > \sigma^2(Y) > \sigma^2(Z)$ we can drop the absolute values and the triangle inequality saturates $\log(\sigma^2(X)/\sigma^2(Y)) + \log(\sigma^2(Y)/\sigma^2(Z)) = \log(\sigma^2(X)/\sigma^2(Z))$. Take however $\sigma^2(X) > \sigma^2(Y)$, $\sigma^2(Y) < \sigma^2(Z)$. In that case the triangle inequality is not satisfied in general, as we have $\log(\sigma^2(X)/\sigma^2(Y)) - \log(\sigma^2(Y)/\sigma^2(Z)) = \log(\sigma^2(X)\sigma^2(Z)/\sigma^2(Y)\sigma^2(Y))$ which in general is not larger or equal than $\log(\sigma^2(X)/\sigma^2(Z))$.

All in all, other proposals that not only are able to quantify ($\mathcal{S}_{1-3}$) but which might rely on more solid mathematical grounds than AVA are needed. An obvious strategy could directly define similarity measures not only from single-point distributions of the *peak* and *pit* sets (mean, variance, etc), but explore higher-order joint distributions of these datasets (two-point, three-point and in general m-point distributions of strings of size m). This however is not likely to be efficient in practice, as high-order statistics usually require access to very long time series (usually the length of the observed time series is

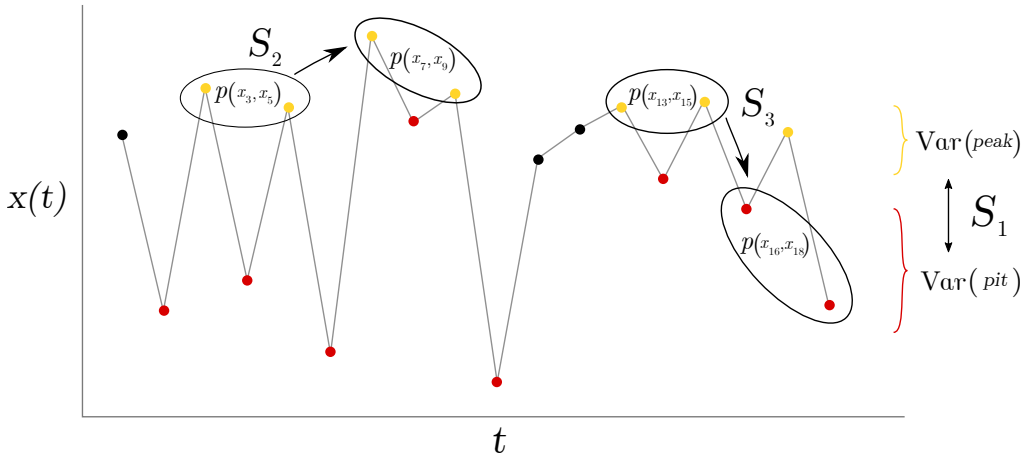


Figure 4.1: An example time series showing the three different scenarios $\mathcal{S}_{1-3}$: $\mathcal{S}_2$ with differing auto-correlations, $\mathcal{S}_3$ with differing cross-correlations between the sets *peak* and *pit*, and $\mathcal{S}_1$ with different marginal distributions (in the case of the example time series here, the sets have different variance (which AVA compares), but in practice we can look at any order moments for any n-point joint distribution).

required to grow exponentially with m).

Furthermore, note that by construction $(peak \cup pit) \subset \mathcal{S}$, meaning that this decomposition is lossy: in a generic fluctuating signal there are data which are neither in the peak nor in the pit set, so a priori important features of the local fluctuations might be lost if one only looks at the peak and pit sets and discard intermediate data. An alternative, which we explore in what follows, is to be able to extract high-order features from peak and pit local neighbourhoods by projecting the full signals into an appropriate topological space.

4.2.2 Visibility graphs and top-bottom VG/HVG asymmetry (ΔVGA)

Recall the definition of the natural visibility graph (Definition 2.2.1 in Section 2.2), and the horizontal visibility graph (Definition 2.2.2 in Section 2.2). Here we label as $G_{\text{top}}(x(t))$ (i.e., the "top" visibility graph) to the graph extracted for either of the two aforementioned procedures (at this point we consider both kinds of graphs VG and HVG separately, although we know that they provide different sorts of information and in practice depending on the particular processes under study it will be more adequate to

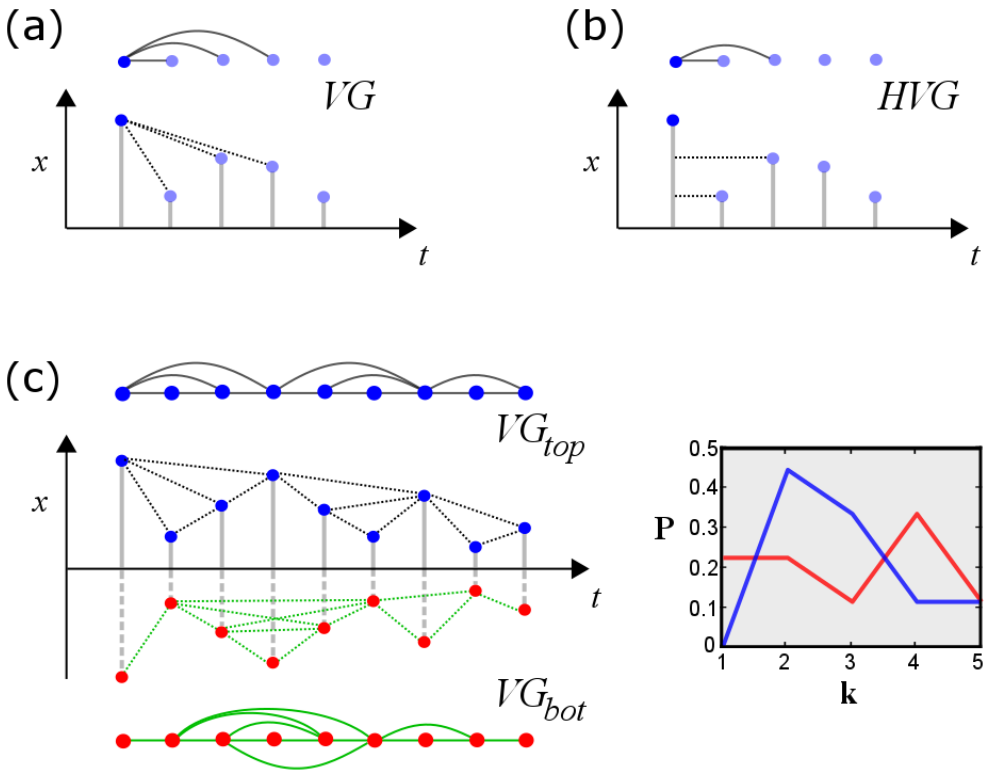


Figure 4.2: (Panels a and b) Sample time series and associated natural visibility (VG, panel a) and horizontal visibility (HVG, panel b) linking criteria. Each mapping is invariant under a certain number of transformations (VG is invariant under affine transformations $x \to ax + b$, $a > 0$, HVG is invariant under monotonic transformations $x \to f(x)$, $f(\cdot)$ an order-preserving function). (Panel c) Sample time series and extraction of the top natural visibility graph (black) and bottom natural visibility graphs, which characterise the visibility structure of local maxima (top) and local minima (bottom) respectively. Note that G_{top} coincides with a standard visibility graph of series $x(t)$, and $G_{\text{bot}}\{x(t)\} \equiv G_{\text{top}}\{-x(t)\}$, that is, characterization of local minima is achieved by extracting the visibility graph from the inverse series $-x(t)$. One can extract features from both G_{top} and G_{bot} (here, a cartoon of the degree distribution $P(k)$) and compare these as a proxy to the comparison between local minima and local maxima statistics.

make use of either HVG or VG). The label "top" comes from the fact that visibility is applied "from above" and therefore tends to encapsulate information on the relative position of local maxima. An obvious drawback of a basic VG/HVG representation is that local minima are hidden, and more specifically the elements in the *pit* set defined above by construction map into nodes with a fixed degree $k = 2$, independently of their actual value. Consider for instance a signal whose odd values build a periodic progression, and whose even values are just noise (see Figure 4.3 for an illustration).

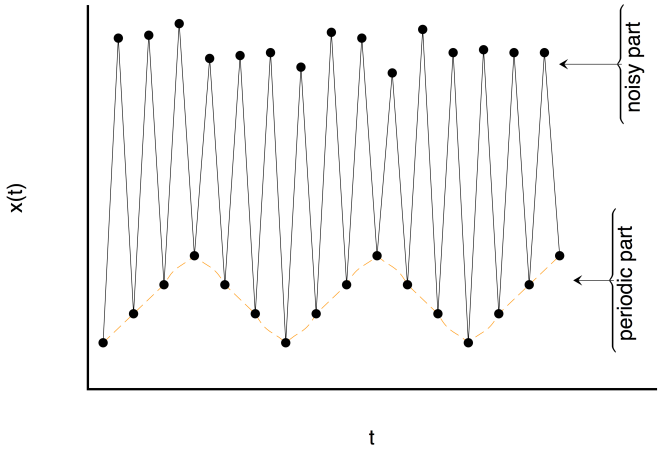


Figure 4.3: Cartoon series having a noisy part at even timesteps and a periodic progression part at odd timesteps: a classical VG/HVG analysis fails to capture the hidden periodic pattern, whereas such pattern can be easily extracted from G_{bot}.

In this simple example, the periodic structure is completely hidden if one only looks at the standard VG/HVG: structure on the *pit* set is lost. As a result, VG/HVG might be insensitive to processes which have two or more spatial scales (For another simple example, consider two time series, the first being a periodic series of period 2, and the second being a mix of two processes: uniform white noise in $[0, 1]$ for even times $t = 2p, p \in \mathbb{N}$, and a constant value > 1 for odd times $t = 2p + 1, p$. The associated VG/HVG is identical despite that both processes are qualitatively different, however using both top and bottom graphs the difference is obvious.). However, this drawback is removed if the VG/HVG algorithm is additionally and subsequently applied "from below". Accordingly, one can also define a "bottom" visibility graph $G_{\text{bot}}(x(t))$ where the visibility criterion is now applied from below, which now will focus *particularly* on the structure of the local minima (highlighting in particular the connectivity structure of the *pit* set) as recently observed [71]. See Figure 4.2 for a sample time series with corresponding G_{top} and $G_{\text{bot}}(x(t))$. Note that this procedure is performed on the whole signal, hence when constructing G_{top} and G_{bot} one is not discarding information on the intermediate data (as it happens with traditional time-domain approaches described above). Furthermore, it is easy to prove that the bottom construction coincides with the top construction if applied on the flipped series, in such a way that the following

identities hold:

$$\forall \mathcal{X} : G_{\text{bot}}(x(t)) = G_{\text{top}}(-x(t)),\ G_{\text{bot}}(-x(t)) = G_{\text{top}}(x(t)). \tag{4.1}$$

Our working hypothesis therefore exploits the potential differences between the top and bottom graphs as a proxy for quantifying the difference between local maxima and local minima statistics in $\mathcal{X}$, or the *top-bottom VG/HVG asymmetry* (ΔVGA) for short. Mathematically, ΔVGA can be quantified in very many different ways (different protocols can be defined in an *ad hoc* way depending on the particular problem under study). For instance, simple *global* topological features that we know are informative include statistics of the degree sequence [72] such as graph's degree distribution $P(k)$, for which, adopting the ℓ_1 norm distance between distributions, leads to a particular definition of the asymmetry $\Delta\text{VGA} = \sum_k |P^{\text{top}}(k) - P^{\text{bot}}(k)|$. This is just an informed choice and in any event, it always comes down to a comparison between certain set of features extracted from the top and bottom VG/HVG. As mentioned above, there are a number of interesting scenarios involving differences in the marginal distributions or correlation structure of peaks and pits ($\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$), and ΔVGA can serve as a tool to investigate these asymmetries. While these issues have not been fully examined in prior work, one important study [71] has used a comparison between G_{top} and G_{bot} with the intention of studying the different dynamics of local minima and maxima in the particular case of sunspot time series. As a matter of fact, sunspot series can be considered the degenerate case because they have a natural absolute and frequently occurring natural minimum (zero), which *by definition* imposes different features on the local minima (including impacting their variance, serial autocorrelation and power spectra). Similar cases where dampening function is applied (if $x > y \Rightarrow x := y$) are similarly unhelpful.

Before presenting the practical algorithmic protocols we explore in synthetic processes the performance of this method, and in particular the ability to outperform current methods as well as the transversality (e.g., how well scenarios $\mathcal{S}_{1-3}$ above are addressed).

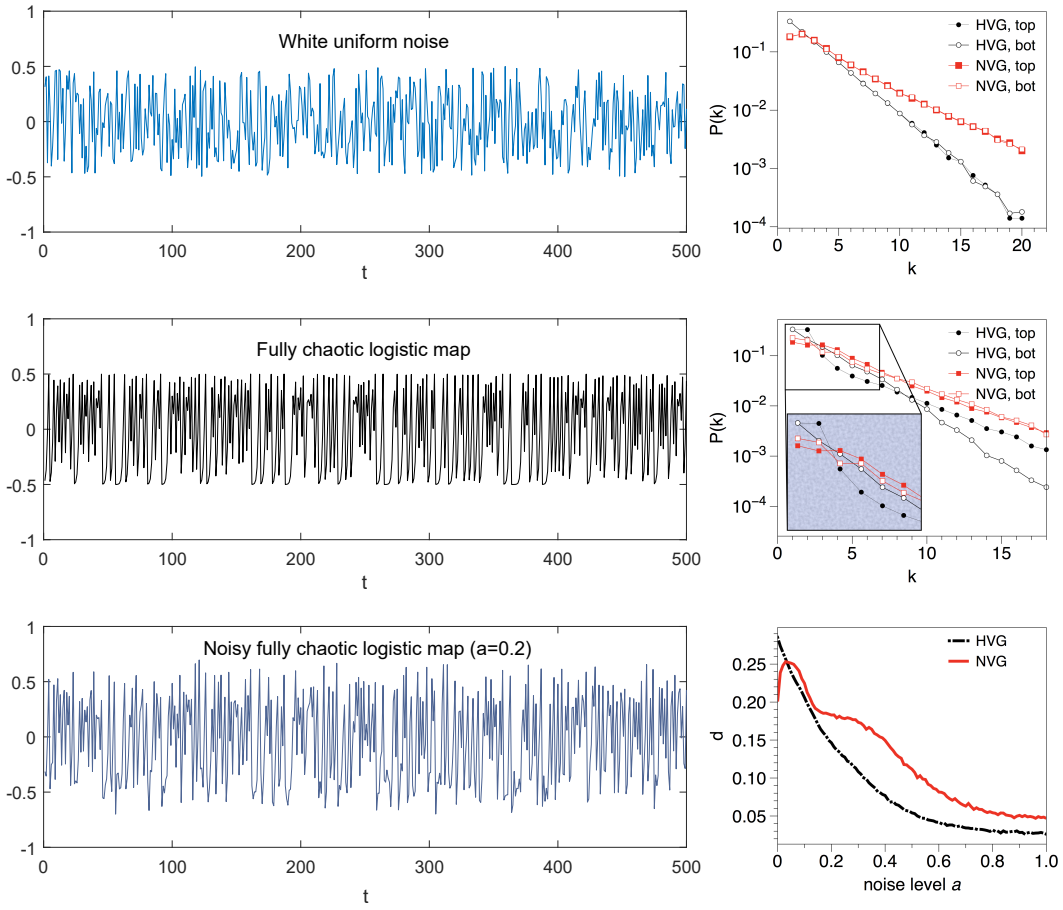


Figure 4.4: (Left panels) Sample time series of uniform white noise (top panel), fully chaotic logistic map (middle panel) and fully chaotic logistic map polluted with a certain amount of white noise (bottom panel). In all cases the signals are irregular and lack any obvious pattern (for appropriate visual comparison all signals have been scaled in the vertical axis). However, whereas the statistics of G_{top} and G_{bot} are equivalent for i.i.d., this is not the case for the chaotic process. In the last case, noise pollutes and hides the chaotic signal and as such the statistics of G_{top} and G_{bot} depend on the noise amplitude. (Right panels) Comparison of the degree distributions extracted from G_{top} and G_{bot} (for both VG/HVG). For the uniform white noise, distributions coincide as expected (the process is invariant under $x(t) \to -x(t)$). For the fully chaotic logistic series, statistics of G_{top} and G_{bot} are different and this feature evidences differences between local minima and maxima. In the last panel we plot the distance between distributions (using ℓ_1 norm) for a fully chaotic logistic map polluted with a certain amount of noise (see the text).

4.2.3 Theoretical validation on synthetic processes

To validate the method, we initially consider a battery of dynamical processes with varying degrees of complexity which focus on different aspects, and consider in every case the performance of a particular definition of $\Delta\text{VGA} = \sum_k |P^{\text{top}}(k) - P^{\text{bot}}(k)|$

(measured on both VG and HVG) and their comparison with the AVA-based metric (VR) (see Figure 4.4 for an illustration). Results are summarized in Table 4. We can highlight the following key results:

1. Symmetric stochastic processes (both with no correlations (white noise) and with rapidly decaying correlations (red noise) yield statistically identical top/bottom VG/HVGs and thus vanishing values of ΔVGA.

2. For non-symmetric white noise, peak and pit statistics are different, even if the process lacks information. In that case, both ΔVGA (applied to VG) and VR detects such asymmetry, while ΔVGA applied to HVG filters out dependences solely based on marginals and predicts no difference: in this sense we conclude that HVG does not capture scenario $\mathcal{S}_1$ as defined above.

   These two initial observations can be summarised in the following theorem:

   Theorem 4.2.1. *Let $\{x(t)\}$ be a bi-infinite time series generated by (i) white noise with a continuous probability density $f(x)$ or by (ii) symmetric correlated red noise. Then $\Delta VGA = 0$ for HVG.*

3. For chaotic processes where peaks and pits are different both in terms of marginal statistics (scenario $\mathcal{S}_1$) as well as in terms of correlation dependences (scenarios $\mathcal{S}_2$ and $\mathcal{S}_3$), all methods successfully detect such asymmetry. In particular, the following theorem holds:

   Theorem 4.2.2. *Let $\{x(t)\}$ be a bi-infinite time series generated by a fully chaotic logistic map. Then $\Delta VGA > 0$ for HVG.*

   A proof for this theorem can be found at the bottom of this section. Gathering together Theorems 4.2.1 and 4.2.2, we have rigorously proved that a chaotic process such as the fully chaotic logistic map can be easily distinguished from white noise, despite the fact that both processes have a flat power spectrum (delta-

distributed autocorrelation function). Additionally, we numerically observe that chaotic processes with a fast-decaying correlation structure are also distinguishable from exponentially decaying correlated noise.

4. Interestingly, there are several cases (such as processes with two alternating dynamics for peaks and pits) where currently used indicators (AVA, AFA) will fail, while ΔVGA efficiently captures significant differences (i.e., VR does not capture scenarios $\mathcal{S}_2$ and $\mathcal{S}_3$ as defined above).

5. ΔVGA is robust against noise pollution and works for short time series, enabling its use in practical cases.

6. VR is a quantity which by construction only depends on the marginal distribution of peaks and pits (in particular, the variance of these marginals), and as such it does not carry information on any temporal correlations, neither intra peaks or intra pits (scenario $\mathcal{S}_2$), nor inter peaks/pits (scenario $\mathcal{S}_3$). It is easy to prove that for a given series $x(t)$, if one breaks down all temporal correlations by reshuffling the series, then the new, reshuffled series $x^{\text{reshuffled}}(t)$ and $x(t)$ will still have the same marginal distributions and thus (under some assumptions) the same VR value, yet $x^{\text{reshuffled}}(t)$ is white noise with a flat spectrum and delta-distributed autocorrelation function, very different in general from the non-reshuffled series $x(t)$. In the same line, VR *breaks down* for any signal which is composed by two alternating processes with similar variances and dynamics with different invariance properties, whereas ΔVGA is not afflicted by these drawbacks.

We now prove Theorems 4.2.1 and 4.2.2.

Proof of Theorem 4.2.1. The proof for this theorem is straightforward. Firstly, we consider the case (i). We rely on a result of [4] where it has been proved that the degree distribution of the HVG of $\{x(t)\}$ generated by white noise is

$$P(k) = (1/3)(2/3)^{k-2} \tag{4.2}$$

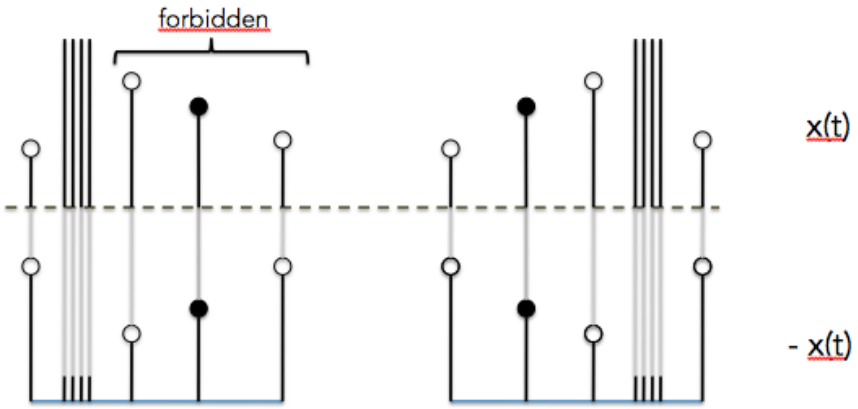


Figure 4.5: Schematic representation of the two sets of diagrams contributing to $P^{\text{bot}}(3)$ (which is equivalent to $P^{\text{top}}(3)$ as computed on $-x(t)$). In every case the reference node x_0 is highlighted as black solid dot. Hidden nodes (an arbitrary large amount of them) are schematised as vertical bars with no dots on top. The first diagram (bottom, left) does not actually appear in $\{-x(t)\}$ as the associated diagram in the original series is forbidden (in the fully chaotic logistic map we never find three consecutive data points in decreasing order [36]). Accordingly, the only set of diagrams is the one pictorially represented in the bottom right. The relative ordering of the data in the original chaotic time series is represented in the upper part.

Trivially, if $\{x(t)\}$ is white noise extracted from $f(x)$, this means that it is a collection of i.i.d. random variables extracted from $f(x)$. But then $\{-x(t)\}$ is again a collection of i.i.d. variables extracted from a different density $g(x)$. By virtue of a theorem proved in [4], the degree distribution of the HVG associated to such series coincides with Equation (4.2), hence $\Delta\text{VGA}= 0$. This closes the first part of the proof.

For (ii), we only need to remark that red noise (AR(1)) is generated via the stochastic map

$$x_t = r \cdot x_{t-1} + \xi, \quad \xi \in N[0,1] \tag{4.3}$$

Trivially, if ξ is drawn from a symmetric distribution (such as the Gaussian function above), then both $\{x(t)\}$ and $\{-x(t)\}$ are equally likely realisations of Equation (4.3), hence ΔVGA vanishes. □

Proof of Theorem 4.2.2. The proof for this theorem is more convoluted than for Theo-

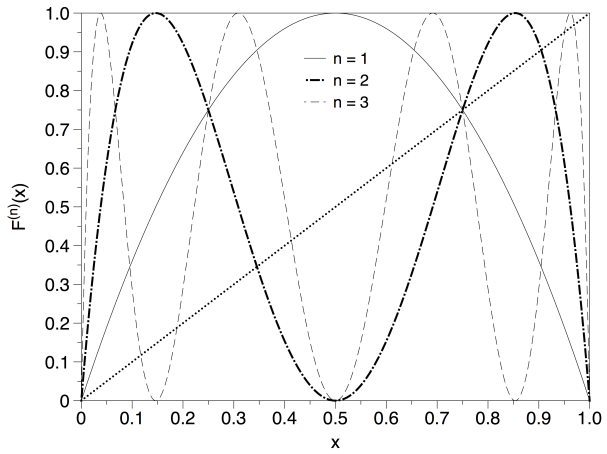


Figure 4.6: A few iterations of the fully chaotic logistic map $x(t+1) = F(x(t));\ F(x) = 4x(1-x)$

rem 4.2.1. Since ΔVGA is semi-positive definite, it vanishes if and only if $\forall k = 2, 3, \cdots$: $P^{\text{top}}(k) = P^{\text{bot}}(k)$. Thus, in order to prove positivity we will only need to find that for a given degree $k = m$, $P^{\text{top}}(m) \neq P^{\text{bot}}(m)$. After a quick numerical inspection, we choose $m = 3$. In [33] it was analytically proved that for a fully chaotic logistic map, $P^{\text{top}}(3) = 1/3$. Without loss of generality, we will prove that $P^{\text{bot}}(3) < 1/3$.

We capitalise on the perturbative expansion formalism advanced in [36] to analytically compute the degree probabilities of HVG associated to chaotic maps with smooth physical invariant measure. This expansion is diagrammatic, and in particular it expands on the number of hidden nodes (see [36] for details). In order to apply this technique here, we first observe that the bottom HVG of $\{x(t)\}$ is equivalent to the top HVG of $\{-x(t)\}$. Accordingly, the set of diagrams that comply with $P^{\text{bot}}(3)$ in the original series is shown in Figure 4.5. It is easy to see that this is an infinitely unfolding combination of diagrams (each with a different number of hidden nodes), and as such formally we have

$$P^{\text{bot}}(3) = \sum_{\alpha=0}^{\infty} P_{\alpha}^{\text{bot}}(3),$$

where α is the number of hidden nodes in each diagram. Let us denote $F(x) = 4x(1-x)$ as the fully chaotic logistic map, $f(x) = \pi^{-1}x^{-1/2}(1-x)^{-1/2}$ as the invariant measure of the map. Since this is a Markovian and deterministic map, the probability of observing

a datum x given that the previously observed datum is y is simply $\delta(x - F(y))$, where $\delta(\cdot)$ is the Dirac-delta distribution. The integral corresponding to each contribution to the probability is then:

$$\begin{aligned} P_{\alpha}^{\mathrm{bot}}(3) = & \int_0^{x_0} f(x_{-1})\delta(x_0 - F(x_{-1}))dx_{-1} \int_0^1 dx_0 \\ & \cdot \int_{x_0}^1 \delta(x_1 - F(x_0))dx_1 \\ & \cdot \left[\prod_{i=1}^{\alpha} \int_{x_1}^1 \delta(z_i - F^{(i+1)}(x_0))dz_i\right] \int_0^{x_0} \delta(x_2 - F^{(\alpha+2)}(x_0))dx_2 \end{aligned} \tag{4.4}$$

The first important point to note is that the first integral can be computed using the scaling properties of the Dirac delta:

$$\int_0^{x_0} f(x_{-1})\delta(x_0 - F(x_{-1}))dx_{-1} = \sum \frac{f(x^*)}{|F'(x^*)|}, \tag{4.5}$$

where x^* are the roots of the equation $F(x) = x_0$ (in particular, we only need to sum up over those roots that belong to the interval of integration, that is, only those roots that satisfy $x^* < x_0$. For $x_0 \in [0, 3/4]$, the only root which fulfils such an inequality is $x^* = (1 - \sqrt{1 - x_0})/2$, hence

$$\begin{aligned} \int_0^{x_0} f(x_{-1})\delta(x_0 - F(x_{-1}))dx_{-1} &= \frac{f(x^*)}{|F'(x^*)|} \\ &= \frac{1}{\pi\sqrt{(1 - \frac{\sqrt{1-x_0}}{2})(1 - \frac{1-\sqrt{1-x_0}}{2})4\sqrt{1-x_0}}} \\ &=: W(x_0). \end{aligned} \tag{4.6}$$

The rest of the integrals only have the effect of reducing the interval of integration of x_0, as these are Dirac integrals and thus are either one or zero, since $\int_a^b \delta(x - y)dx$ is one if $y \in [a, b]$ and zero otherwise.

Let us assume for a moment that we remove in the integral above any contribution of nodes above x_1. After a little algebra, this particular case labelled Q would yield a

probability

$$\begin{aligned} Q &= \int_0^{x_0} f(x_{-1})\delta(x_0 - F(x_{-1}))dx_{-1} \int_0^1 dx_0 \int_{x_0}^1 \delta(x_1 - F(x_0))dx_1 \\ &= \int_0^{3/4} \frac{f(x^*)dx_0}{4\sqrt{1-x_0}} \\ &= \int_0^{3/4} W(x_0)dx_0 \\ &= 1/3 \\ &= P^{\text{top}}. \end{aligned} \tag{4.7}$$

where $W(x_0)$ was defined in Equation (4.6), and the shrinking of the interval of integration in x_0 from $[0,1]$ to $[0,3/4]$ comes from the integral $\int_{x_0}^1 \delta(x_1 - F(x_0))dx_1$, which is only equal to one if $F(x_0) > x_0$, which happens for $x_0 \in [0,3/4]$ (see Figure 4.6). Importantly, any other additional Dirac delta integral (associated to the contribution of further nodes, hidden or otherwise) of the form $\int_a^b \delta(x-y)dx$ will have the effect of *further shrinking* the interval $[0,3/4]$ in the integral over x_0, in such a way that for each order α the original interval of integration of x_0 shrinks into a subinterval $\mathcal{I}_\alpha$. To showcase such situation let us consider in full detail the computation of $P_0^{\text{bot}}(3)$ and $P_1^{\text{bot}}(3)$.

$\underline{\alpha = 0.}$

$$\begin{aligned} P_0^{\text{bot}}(3) = &\int_0^{x_0} f(x_{-1})\delta(x_0 - F(x_{-1}))dx_{-1} \int_0^1 dx_0 \\ &\cdot \int_{x_0}^1 \delta(x_1 - F(x_0))dx_1 \int_0^{x_0} \delta(x_2 - F^{(2)}(x_0))dx_2 \end{aligned} \tag{4.8}$$

Let's consider the last two integral terms above: they require $F(x_0) > x_0$ and $F^{(2)}(x_0) < x_0$. Looking at Figure 4.6 and considering the fixed point structure of $F(x)$ and higher iterates of the map, we find that the first condition shrinks the interval of integration of x_0 from $[0,1]$ to $[0,3/4]$, whereas the second condition is fulfilled for $x_0 \in [\frac{5-\sqrt{5}}{8}, \frac{3}{4}]$. The intersection of the two conditions above yields an interval

$$\mathcal{I}_0 = [\frac{5-\sqrt{5}}{8}, \frac{3}{4}],$$

and therefore

$$P_0^{\text{bot}}(3) = \int_{\frac{5-\sqrt{5}}{8}}^{3/4} W(x_0)dx_0 \approx 0.1333$$

$\underline{\alpha = 1.}$

$$\begin{aligned} P_1^{\text{bot}}(3) = & \int_0^{x_0} f(x_{-1})\delta(x_0 - F(x_{-1}))dx_{-1} \int_0^1 dx_0 \\ & \cdot \int_{x_0}^1 \delta(x_1 - F(x_0))dx_1 \\ & \cdot \int_{x_1}^1 \delta(z_1 - F^{(2)}(x_0))dz_1 \\ & \cdot \int_0^{x_0} \delta(x_2 - F^{(3)}(x_0))dx_2 \end{aligned} \tag{4.9}$$

Let's then consider the last *three* integral terms above: they require

$$F(x_0) > x_0, F^{(2)}(x_0) > F(x_0), F^{(3)}(x_0) < x_0.$$

Looking at Figure 4.6 and considering again the fixed point structure of $F(x)$ and higher iterates of the map, we find that the first condition shrinks the interval of integration of x_0 from $[0, 1]$ to $[0, 3/4]$, whereas the second condition is fulfilled now for $x_0 \in [0, 1/4] \cup [3/4, 1]$. The third condition is fulfilled in $x_0 \in [0.11698, 0.188255] \cup [0.41318, 0.61126] \cup [3/4, 0.95048] \cup [0.96985, 1]$. The intersection of all these conditions thus provides

$$\mathcal{I}_1 = [0.11698, 0.188255],$$

and therefore

$$P_1^{\text{bot}}(3) = \int_{0.11698}^{0.188255} W(x_0)dx_0 \approx 0.03174$$

One can proceed indefinitely adding higher orders in α, which yield smaller and smaller contributions to the total probability. In other words, for any $\alpha > 0$ the resulting interval $\mathcal{I}_\alpha$ is always a subset of $[0, 3/4]$. Since the integrand in the last integral of Equation (4.7) is always positive (see Figure 4.7 for a graphical support), we conclude that $P_\alpha^{\text{bot}}(3) < Q$. Furthermore, we need to prove that if we concatenate the subinterval obtained for each

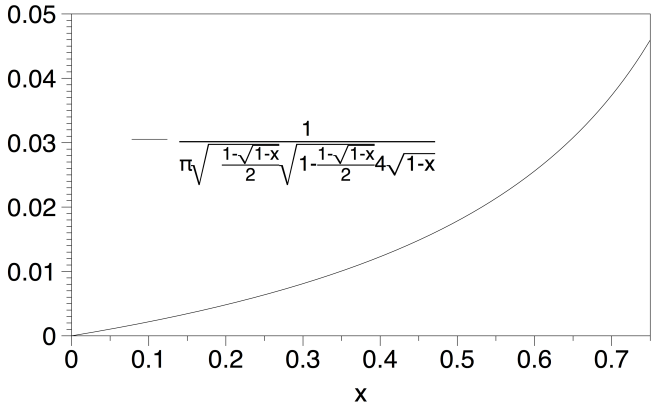


Figure 4.7: Integrand in the last integral of Equation (4.7).

α, its union will always be smaller than $[0, 3/4]$:

$$\bigcup_{\alpha=0}^{\infty} \mathcal{I}_\alpha \subset [0, 3/4] \tag{4.10}$$

This last statement can be easily proved by showing that there are always subintervals of $[0, 3/4]$ not covered in any $\mathcal{I}_\alpha$. In particular, consider the subinterval $[1/4, (5-\sqrt{5})/8]$. We have explicitly shown that such subinterval is not included in $\mathcal{I}_0$, nor in $\mathcal{I}_1$. Now, for $\alpha \geq 1$ by construction one of the conditions which is always present is $F^2(x_0) > F(x_0)$. Now, such condition is fulfilled for $x_0 \in [0, 1/4] \cup [3/4, 1]$, leaving our interval $[1/4, (5-\sqrt{5})/8]$ out. As this situation holds $\forall\alpha$, this directly yields Equation (4.10). Finally, since integration is monotonic and since the integrand $W(x_0)$ is positive, we trivially have

$$\mathcal{A} \subset \mathcal{B} \Rightarrow \int_{\mathcal{A}} W(x_0)dx_0 < \int_{\mathcal{B}} W(x_0)dx_0.$$

Altogether, this means that $P^{\text{bot}}(3) = \sum_{\alpha=0}^{\infty} P_\alpha^{\text{bot}}(3) < Q = P^{\text{top}}(3)$, which concludes the proof.

□

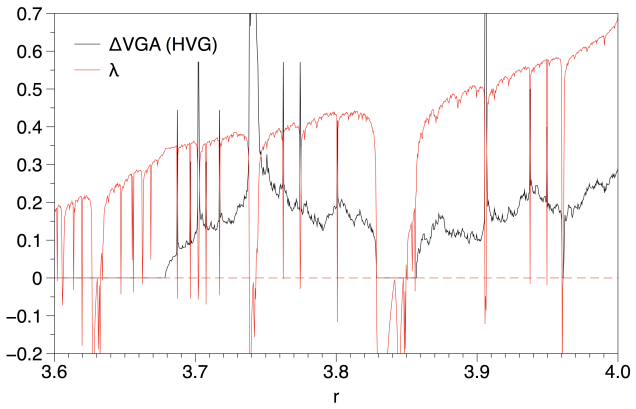


Figure 4.8: Plot of ΔVGA (defined as the ℓ_1 distance between top and bottom degree distributions) for time series extracted from the logistic map $x(t+1) = \mu x(t)[1-x(t)]$, for a range of values of μ for which the map generates chaotic trajectories of different type, interspersed by periodic windows. In the same figure we plot the Lyapunov exponent of the map.

4.2.4 Parametric Analysis

Finally, in order to explore the relationship between ΔVGA and a number of standard linear and non-linear indicators (mean, variance, skewness, kurtosis, Lyapunov exponent), we have considered a parametric map (the logistic map) $x(t+1) = \mu x(t)[1-x(t)]$. As μ smoothly varies between 3.6 and 4 the map generates chaotic (aperiodic) time series with different degree of chaoticity (different Lyapunov exponent, attractor's fractal dimension and physical invariant measure), interspersed by periodic windows. We recorded each statistic as we continuously scan μ. In Figure 4.8 we plot ΔVGA (defined as the ℓ_1 distance between top and bottom degree distributions for time series) extracted from the logistic map $x(t+1) = \mu x(t)[1-x(t)]$, for a range of values of μ for which the map generates chaotic trajectories of different type, interspersed by periodic windows. In the same figure we plot the Lyapunov exponent of the map. Both indicators are only weakly correlated (Pearson correlation coefficient $r = 0.22$). Correlation between ΔVGA and other standard indicators include: $r = -0.104$ (mean), $r = 0.2$ (standard deviation), $r = -0.12$ (skewness), $r = 0.158$ (kurtosis). We conclude that ΔVGA finds negligible correlation with all indicators, except for standard deviation and Lyapunov exponent, where the correlation was found to be weak: ΔVGA provides genuinely different infor-

mation than standard linear and nonlinear indicators. Additionally, note that ΔVGA is notably unrelated to power spectrum (for instance, ΔVGA is positive for $\mu = 4$ and null for white noise, and both processes have identical power spectrum).

From the theoretical analysis above, we conclude that ΔVGA as defined above can be efficiently used to detect and quantify subtle differences between the statistics of local maxima and local minima in the associated series which are not correlated with standard linear and non-linear indicators. If such asymmetry is solely based on the (one-point) marginal distributions of peaks and pits (scenario $\mathcal{S}_1$), then all three choices (ΔVGA applied to VG and HVG, and AVA) are qualitatively similar. However, in the event the peaks and pits have identical marginals but different correlation structure, then only ΔVGA capture this difference. Finally, if there is a difference between marginals but no difference in the correlation structure, then ΔVGA (applied to VG) and AVA accurately capture the difference, but not ΔVGA applied to HVG (as this latter is an order statistic).

4.3 Data extraction methods and visibility protocol

We have applied the methodology in three different situations: EEG and fMRI data, financial time series of NYSE (the same data we used in Chapter 3), and worldwide temperature records. In what follows we provide details on data acquisition and pre-processing, as well as the protocols devised in each case to compute ΔVGA.

4.3.1 EEG-fMRI data

The functional neuroimaging data that we analysed for the current study were collected in a multi-modal acquisition that also included EEG data. Because all EEG and fMRI pre-processing, artifact correction, and fMRI series construction details for this data set have been reported previously [67], we do not report the details again. In brief, the fMRI data were acquired during wakeful rest and sleep from 63 participants, of which at least 55 reached N1 sleep. EEG data were acquired contemporaneously, filtered, and here

Process	**Definition**	**Expected outcome**	ΔVGA of : VG	HVG	AVA
(symmetric) uniform white noise	$x(t) = \xi$, $\xi \sim U[-1, 1]$	0	✓	✓	✓
exponentially correlated red noise	$x_t = rx_{t-1} + \xi, \quad \xi \in N(0, 1)$	0	✓	✓	✓
Power law (non-symmetric) white noise	$x(t) = \xi$, $\xi \sim x^{-\alpha}$	0	✓	×	×
Fully chaotic logistic map	$x_t = 4x_{t-1}(1 - x_{t-1})$	> 0	✓	✓	✓
Fully chaotic logistic map polluted with white noise	$\rho_a(t) = x(t) + \xi$, $\xi(t) \sim \mathcal{U}(-a, a)$, $a \in [0, 1]$, $x(t)$ fully chaotic logistic map.	> 0 but decreasing with noise amplitude	✓	✓	✓
Selected pollution	$y_t = x_t + \xi$, x_t fully chaotic logistic map, $\xi = \begin{cases} \sim U[0, a], & x_t < 0.3 \\ 0 \quad \text{otherwise} \end{cases}$,	> 0 but decreasing with noise amplitude	✓	✓	✓
Alternating dynamics #1: x_t fully chaotic logistic, $\xi \sim U[1 - a, a]$	$y_t = \begin{cases} \sqrt{8}ax_t, & t \text{ even} \\ \sqrt{12}\xi \quad \text{otherwise} \end{cases}$	> 0	✓	✓	×
Alternating dynamics #2: x_t fully chaotic logistic, $\xi \sim U[1 - a, a]$	$Y_t = \begin{cases} Ca \cdot \sin(\pi t/20), t \text{ even} \\ \sqrt{12}\xi \quad \text{otherwise} \end{cases}$ where C is such that the variance of the sinusoidal process is one	> 0 and independent of a until both processes coalesce ($a > 0.5$)	✓	✓	×

Table 4: Summary of results on synthetic processes. In every case we extract time series from the dynamical equations defined on the first column, and compare the results provided by the statistics ΔVGA (applied to VG and HVG) and AVA (computed from the logarithm of the Variance Ratio), see the text for details.

used solely for determining state of awareness (sleep staging: wakefulness, N1, N2 and N3 sleep). It is important to note that physiological noise correction [73] was applied to the fMRI data (for details, see [67]) and those data were also de-spiked prior to analysis. All methods were performed in accordance with the relevant guidelines and regulations, and informed consent for participation was provided by all participants in the fMRI/EEG study.

4.3.2 Calculating ΔVGA and protocols

4.3.2.1 Neurobiological case

In a neurobiological context we were interested in monitoring systematic differences between regions of the degree distributions (rather than a net value out of the comparison of the whole distributions) which hold across participants in specific brain areas. A scalar projection such as the one defined for ΔVGA in the synthetic cases lose such fine-grained level of detail. Therefore, in this particular application we first constructed, for each voxel, degree distributions up to degree $k = 9$ (this histogram truncation is justified as the modal degree value in fMRI series is around 3 or 4 with very few time points having a degree greater than 8, as shown in the results section). A second reason for selecting $k = 10$ as a limit was our interest in local features of the time series, rather than in any high-amplitude but infrequent spikes isolated from each other by more than 22 (TR = $2.2s \times 10$ samples) seconds. We recall that Figure 4.2 illustrates graphically the concept of how node degrees are established, and how a time series can be characterized via degree distributions of the top and bottom VG/HVGs. As an additional pre-processing, in this application we created empirical cumulative distribution (CDF) functions of each histogram, normalizing the bar heights so that the area of the histogram is equal to one.

4.3.2.2 Normalization to common spatial reference space

After creating maps of visibility graph differences (top-bottom) on a single voxel level, we obtained a transformation between each EPI time series to its corresponding anatomical image using FSL's epi-reg script. The most important steps in this procedure are

FAST's (FMRIB's Automated Segmentation Tool; [74] histogram based segmentation of the T1 structural scans to derive white matter maps, and the use of the boundaries of these white matter maps to perform Boundary-Based co-Registration of the EPIs to their corresponding T1 structural images (BBR; [75]). We then performed nonlinear normalization (FNIRT), of each subject's T1 images into $2 \times 2 \times 2$ MNI space. Importantly, we concatenated the two transformations (EPI to T1; Linear, guided by white matter boundaries) and T1 to MNI (nonlinear warp) to derive a single transformation from EPI space to MNI space. We used this transformation to align the AVA maps from original space to the MNI template in a single step.

4.3.2.3 Single-voxel calculations, and group-level analyses

The sleep-staging procedure allowed us to obtain fMRI time series for wakefulness (W) and three different sleep stages: N1, N2 and N3. On the single participant level, for each of the four conditions, we conducted voxel-wise analyses to derive the node-degree histograms for the top and bottom graphs (i.e., these were derived for each voxel). These were represented as the empirical cumulative distribution function. With these CDFs we could answer the following two questions:

1. On the single-condition level we identified brain areas that differentiated the top from bottom CDFs; we refer to this 'difference' histogram as *Asymmetry of Visibility Histogram* (AoVH), which is our primary way of implementing ΔVGA in this application. We performed these analyses for CDFs derived from both VG and HVG. It is important to note that because these CDFs were normalized, a main effect of condition is not possible (the histograms average across bins to one) and there will always be a strong main effect of bins. Instead, it is the interaction between top/bottom analysis and Bin which we use to show differences across conditions. In reference to the illustration in Figure 4.2 (panel C), we are interested in whether the difference between the blue and red histograms, calculated for a given voxel's time series' are systematic across participants.

2. Then, in order to study differences between conditions, we compared the AoVH of the two conditions. We did this by deriving AoVH for each condition and then determining statistically whether these differed between the two conditions. Note that comparisons between conditions were based on time series matched for length within participant.

On the group level, to evaluate voxels showing statistically-significant AoVH *within* each study condition (W, N1, N2, N3) we conducted voxel-wise repeated measures ANOVA with two fixed factors; Time series view (two levels [Top, Bot]), and node-degree-histogram Bin (8 levels as explained above). To enable inferences about the population, we modelled participants as a random factor. In order to compare *between* any two conditions (e.g., W vs. N1) we derived the AoVH for each condition, and conducted a repeated measures ANOVA with 2 fixed factors: Condition (e.g., W vs. N1) and Node degree histogram bin of the difference histogram (8 bins, values 2(min)-9(max)). We note that 8 bins are used here, as the 9th bin would not be independent of the prior 8 (due to normalization).

These analyses returned a statistical significance value for the interaction term, for each brain voxel. We then implemented a clustering procedure [76] to identify brain areas where many contiguous voxels, each with a p value of $\leq .001$ show a significant interaction: this identifies an "activation cluster".

4.3.2.4 Financial case

In application to financial data, we use the same data that we introduced in Chapter 3: a dataset of financial stocks comprising stock evolution of 35 major American companies from the New York Stock Exchange.

In previous sections we have confirmed that the (ℓ_1) distance between top and bottom degree distributions of VG/HVG is an efficient and informative definition of ΔVGA which in many cases outperforms existing indicators for peak/pit asymmetries and is genuinely different from standard linear and nonlinear indicators. This is precisely the

scalar which we shall use in the financial context. Our protocol is as follows: we compute the ΔVGA by splitting the time series for each company into yearly time series and calculate the distance $d = \sum_k |P^{\text{top}}(k) - P^{\text{bot}}(k)|$ over the associated HVGs for each year. Accordingly, each year is characterized by a vector of distances. For instance, for year 1998, we have a vector of 35 dimensions whose entry i is the ΔVGA for company i in 1998. Subsequently, a principal component analysis is performed to dimensionally reduce data, and a projection into the two first principal components is used to visually cluster different years. This is a similar method to the one we used in Section 3.3.3.

4.3.2.5 Climatic case

This application is conceptually similar to that reported for financial time series. Rather than considering stock prices of major shares in multiple years, we considered daily temperature data sampled over a global grid with a resolution of 192 longitude and 94 latitude, with average temperature for each day (365 values) provided for each grid point (data obtained from [77]). For each year between 1995 and 2015 we compute ΔVGA following the same definition used in the financial setting.

4.4 Results

4.4.1 Application to fMRI

We first compared the top and bottom visibility graphs *within* the Wakefulness and the three sleep stages (N1, N2 and N3). As shown in Figure 4.9, we found significant differences, predominantly in thalamic and frontal regions, for each of these conditions, with the exception of the case of VG in the N3 sleep condition. To better understand these results, we determined which visibility values tended to contribute most strongly to the statistically-significant differences in degree distributions that produced the Wakefulness (W) results in Figure 4.9 (rows 1, 2). To this end, for each cluster we derived a histogram that communicated the visibility bins that most strongly differentiated the top and down degree distributions for each voxel in the cluster. We did this by (i) transforming the

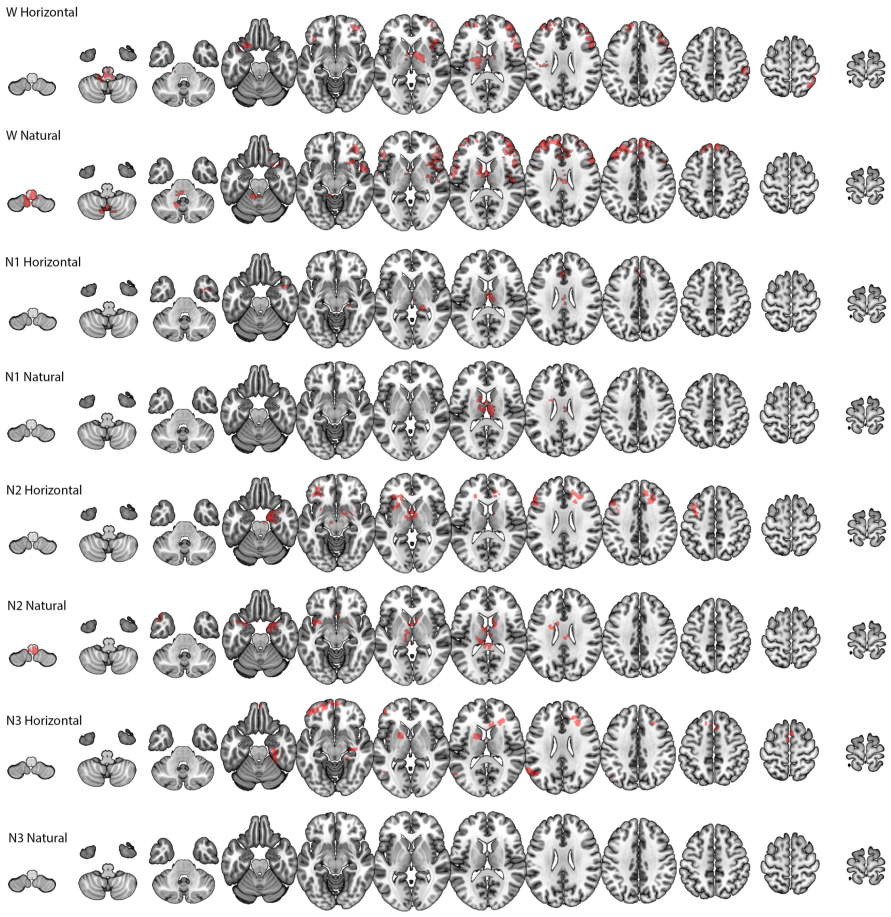


Figure 4.9: For wakefulness (W) and each sleep stage (N1, N2, N3), the figure shows brain regions for which the degree distributions of the top and bottom visibility graphs differed significantly as determined by an ANOVA repeated-measures analysis across participants. Horizontal=derivation from HVG; Natural=derivation from VG. Images were generated using Neuroimaging FSL software.

cluster to original space, (ii) for each voxel, identifying the bin that maximally differentiated the top from down histogram, (iii) transforming that value to common space, (iv) creating an average across participants for each voxel in the cluster. The resulting histograms communicated a very clear and consistent result: for VG the modal degree value that maximally differentiated the top and bottom histograms was 4, with narrow tails towards the values of 3 and 5 (indicating that for some participants, some voxels maximally differentiated the histograms at values 3 and 5). Importantly, there were no cases with means below 3 or above 5. For HVGs, in *all* clusters the modal degree value that maximally differentiated the top and bottom histograms was 3, with very narrow

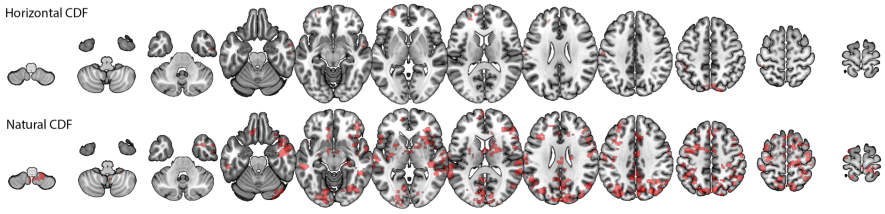


Figure 4.10: Brain areas where spontaneous activity patterns differed between wakefulness and N2 sleep. In these areas, the difference between top and bottom graphs varied systematically between wakefulness and sleep. Images were generated using Neuroimaging FSL software.

tails towards 2 and 4.

These findings are very important as they show that the differences identified by ΔVGA (AoVH) were driven by very local dynamics rather than due to differences in propensity of isolated extreme events.

In a separate analysis, we also found that ΔVGA profiles could discriminate wakefulness from sleep. As shown in Figure 4.10, we identified numerous areas, in both occipital (visual cortex) and lateral temporal cortices, where dynamics during wakefulness and N2 sleep differed significantly. No differences were found between W and N1 or between W and N3.

4.4.2 Application: unsupervised clustering of financial periods

First, we show in the upper panels of Figure 4.11 that ΔVGA is not correlated with the measure based on Variance Ratio (VR, left panel, Pearson correlation $r = 0.085$), nor with the standard econometric measure to capture financial series fluctuations (annualized volatility, right panel, Pearson correlation $r = 0.032$). This means that ΔVGA is a genuinely different measure in this domain. The correlation between an index and its volatility is related to the skewness of the index, hence this result further validates the hypothesis that ΔVGA is independent of skewness.

Then, as we did in Section 3.3.3, we perform principal component analysis (PCA) on

the vector $\{d^{\mathbf{c}}(\text{year})\}_{\text{year}=1998}^{2012}$ where $d^{\mathbf{c}}(\text{year})$ is the ΔVGA (HVG) for company $\mathbf{c}$ for a specified year, and project results on a two-dimensional space spanned by the first two principal vectors of the PCA projection. This plot is shown in the left panel of Figure 4.11, and indicates that the measure based on ΔVGA is informative and we can cluster periods of financial turmoil together (the global financial crisis 2009-2010 is found separated from the dot-com bubble 1998-2000 and from the rest of the years). In the right panel of the same figure, we further compare equivalent plots produced via $|\log(VR)|$. We thus compute $|\log(VR)|$ for each year, for each company, and perform PCA on the space $\{|\log(VR)|^{\mathbf{c}}_{(\text{year})}\}_{\text{year}=1998}^{2012}$. In this case no robust clustering appears.

4.4.3 Application: global daily temperature time series

The ΔVGA map for the year 2015 is seen in Panel A of Figure 4.12. As a first step we quantified the relation between ΔVGA and several moments of the temperature distribution. To this end, we derived ΔVGA spatial heat maps for each of the years 1995-2015, and we then calculated the correlation between the observed value of ΔVGA (per grid point) and the following parameters of the yearly temperature time series: (i) mean, (ii) standard deviation, (iii) skewness , and (iv) kurtosis. Note that this returns a set of pair-wise correlation values *per year* and we can then examine the distribution of these correlation values across years. Similarly to what we found for the validation case, the correlation with mean was low (Mean Pearson's R; $r = .06 \pm .03$; the correlation with standard deviation, r = 0.24 $\pm$.05; with skewness, r = -0.09, $\pm$0.02 and with kurtosis, r = -0.13, $\pm$0.04. Note the low correlation with skewness, which strongly suggests that ΔVGA measure is not loading on extreme events [70]. The correlation between ΔVGA and the standard deviation was moderate, and the correlations with mean and kurtosis minimal. This suggests that ΔVGA provides information not captured by typical moments.)

A Principal Component Analysis applied to the 21 Distance maps (1995-2015) identified a first component that accounted for 40% of the variance, and a second that accounted

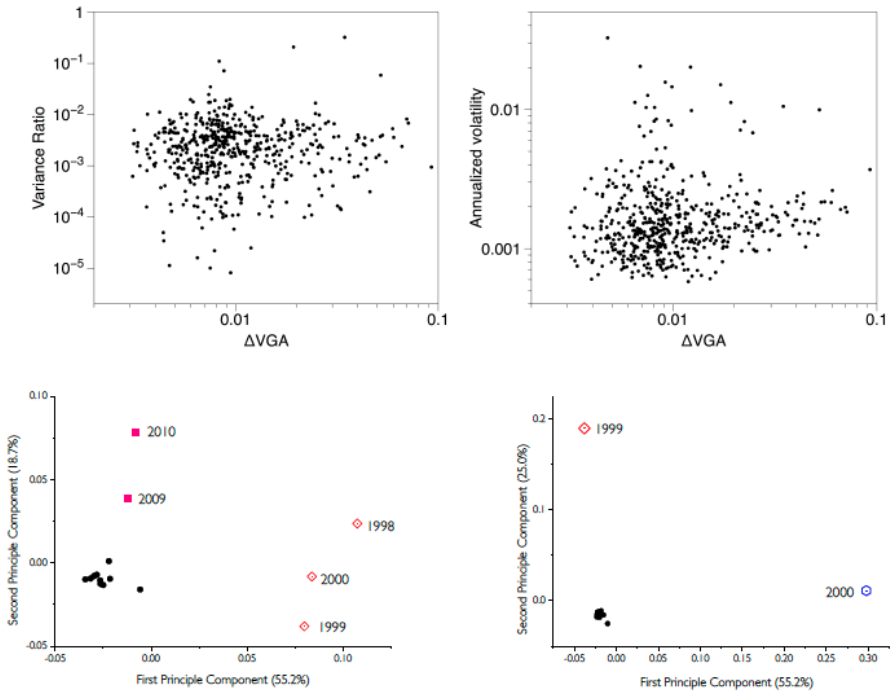


Figure 4.11: (Upper panels) Scatter plot of the ΔVGA (HVG) versus Variance Ratio (left panel) and annualized volatility (right panel) for each year in the period 1998-2012 for the set of 35 companies in the dataset ($N = 525$ samples). No obvious correlation shows up (Pearson correlation $r = 0.085$ and 0.032 respectively). (Bottom, Left panel) PCA on the space $\{d^{\mathbf{c}}(\text{year})\}_{\text{year}=1998}^{2012}$ where d is the distance between the top and bottom degree distributions (each point is a 35-dimensional vector describing the financial state of a given year in the PCA space spanned by the distance of top vs bottom degree distributions, for each of the 35 stock prices). Here we project points in the 2-dimensional space spanned by the first two principal components. We can see how three clusters emerge naturally: one agglomerating the years 1998,1999,2000 (which coincide with the .com bubble), one agglomerating 2009,2010 (coinciding with the period of the global financial crisis), and another cluster with the rest of the years (which coincide with periods of relatively stable financial activity). (Bottom, Right panel) Equivalent PCA projection on the space $\{|\log(VR)|^{\mathbf{c}}_{(\text{year})}\}_{\text{year}=1998}^{2012}$, where the feature extracted from each stock price is now $|\log(VR)|$. We find that this feature is less informative. Accordingly, we therefore interpret that the difference between local maxima and local minima are not simply related to differences between marginal distributions, but a difference in the correlation (temporal ordering) structure.

for 5% (see Figure 4.12). The yearly loadings on the first component were quite stable across the years. In contrast, the yearly loadings on the second component show a strong linear change with time.

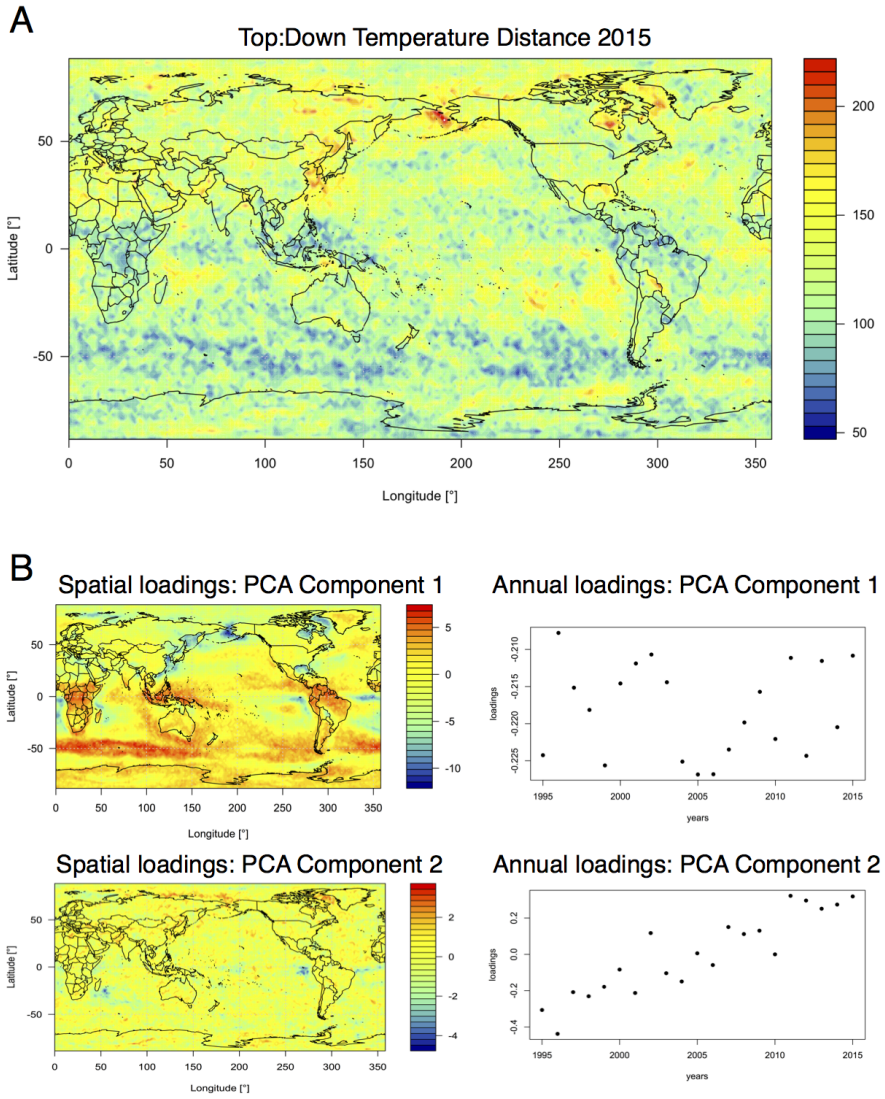


Figure 4.12: Panel A: Distribution of Top:Down Distance values calculated from yearly temperature time series in 2015, with warmer colours indicating greater difference between top and bottom visibility graphs. Panel B: results of PCA applies to Top:Down Distance maps for the years 1995-2015. The first component accounted for around 40% of the variance with stable yearly loadings. Component two accounted for around 5% of the variance and showed marked changes in yearly loadings over time. Images were built using R software [78] with ncdf4 package by Uri Hasson, code available under request.

Chapter 5

The spectral properties of visibility graphs: Feigenbaum graphs

5.1 Introduction

In the context of visibility and horizontal visibility graphs, a great deal of effort has been paid to study properties of these graphs based on their degree sequence [36, 72], but much less attention has been directed to their spectral properties. Nevertheless, the *Graph Index Complexity* (GIC) [79], a rescaled quantity of the maximal eigenvalue of the graph's adjacency matrix, has been proposed as a measure to characterise the complexity of the associated sequence, and has been used in several applications including detection of Alzheimer's disease [30] or epilepsy [80] among others [81, 82], or even the discrimination between randomness and chaos [83].

However, at odds with the case of the degree sequence, a basic theoretical understanding of the spectral properties of HVGs is still lacking. This is the main aim of this chapter. To achieve this aim, we generate sequences (trajectories) from the logistic map, as this

is a well-known map which generates both periodic and chaotic sequences, allowing us to explore spectral properties of HVGs associated to different classes of time series. In previous works, the HVG of a time series generated by the logistic map for a specific value of the parameter μ was coined as a *Feigenbaum graph* [84]. In a nutshell, Feigenbaum graphs are HVGs associated with the Feigenbaum scenario, where one-dimensional unimodal maps exhibit a period-doubling route to chaos. In other words, Feigenbaum graphs are a special subset of HVGs. In this chapter we give a first look at some spectral properties of this family of graphs.

The rest of the chapter goes as follows. In Section 5.2 we define and provide a basic characterisation of Feigenbaum graphs below and above the accumulation point. We show that below the accumulation point, Feigenbaum graphs are easily enumerable in terms of a two-parameter family of graphs which can be generated in terms of two graph operations and admit an algebraic structure. Such enumeration is not possible above the accumulation point, where we show that for particular values of the map's parameter we no longer have unique graphs but an ensemble of them. In Sections 5.3 and 5.4 we explore the spectral properties including the spectrum of the adjacency matrix (with special interest in the largest eigenvalue) above and below the accumulation point. For the chaotic region, we finally compare the results to those associated with an i.i.d process. In Section 5.5 we conclude.

5.2 From Horizontal visibility graphs to Feigenbaum Graphs

Recall the definition of the natural visibility graph (Definition 2.2.1 in Section 2.2), and the horizontal visibility graph (Definition 2.2.2 in Section 2.2). In addition, we note that the mean degree $\bar{d}$ of the HVG associated to a periodic series with period T is [84]

$$\bar{d} = 4\left(1 - \frac{1}{2T}\right) \tag{5.1}$$

Here we consider the set of HVGs generated from trajectories of the well-known logistic map $x_{t+1} = \mu x_t(1-x_t)$ where $\mu \in [0, 4]$ is a parameter and $x_t \in [0, 1]$. This is a unimodal

map that undergoes a period-doubling bifurcation route to chaos as the parameter μ is increased. For $\mu < \mu_\infty \approx 3.569...$ the attracting set consists of periodic orbits with period $T = 2^n$, where n is an integer that increases without bounds as μ approaches the accumulation point μ_∞. For any given integer $n \geq 0$, one can associate a range of values $I_n = [\mu_n, \mu_{n+1})$ where μ_n is the value of the map's parameter for which a stable periodic orbit of period 2^n first appears (with $\lim_{n\to\infty} \mu_n \equiv \mu_\infty$), whereas at μ_{n+1} the 2^n-periodic attractor becomes unstable and a 2^{n-1}-periodic attractor emerges via period-doubling bifurcations. Therefore, for a given value of $\mu < \mu_\infty$, there exists a single stable attractor and, after a transient, any initial condition converges into the same periodic time series.

HVGs generated from trajectories of the logistic map have been studied before, and have been coined as *Feigenbaum graphs* [84]. We start by formally introducing these:

Definition 5.2.1 (The Infinite Feigenbaum graph)**.** Consider a periodic orbit of period $T = 2^n$ from the logistic map, and build a time series of $2NT + 1$ data (with $N \in \mathbb{N}$), as

$$\mathcal{S} = \{x_{-NT}, x_{-NT+1}, \ldots, x_{-1}, x_0, x_1, \ldots, x_{NT-1}, x_{NT}\}.$$

The associated HVG is referred to as a Feigenbaum graph [84]. In the limit $N \to \infty$, the associated HVG is a locally finite infinite graph. It is denoted F_n^∞ and is referred to as the infinite Feigenbaum graph [84].

Remark. As the infinite Feigenbaum graph is a connected, locally finite, infinite graph, it is countable [85]. Since for a fixed value n all values of $\mu \in I_n$ generate, once the transient is discarded, the same periodic time series, then it is obvious that F_n^∞ is unique.

A sketch of F_n^∞ for a few values of n is depicted in Figure 5.1.

5.2.1 Feigenbaum graphs with $\mu < \mu_\infty$: a simple parametrisation F_n^k

Observe that for a fixed $n < \infty$ (that is, for $\mu < \mu_\infty$), the trajectory generated by the logistic map is (after an irrelevant transient) the same periodic series of period $T = 2^n$.

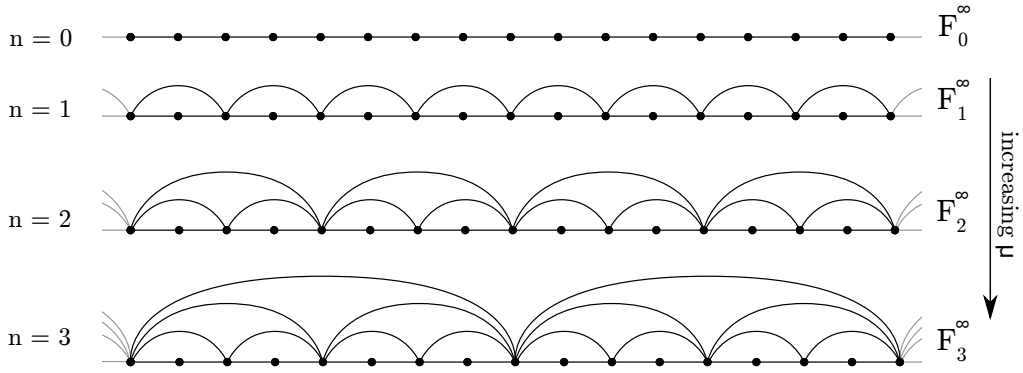


Figure 5.1: Sketch of the family of (infinite) Feigenbaum graphs F_n^∞ for $\mu < \mu_{\infty}$., displaying the sequence of graphs associated to the periodic attractors of increasing period $T = 2^n$ of an unimodal map undergoing a period doubling cascade.

In these cases, the Feigenbaum graph is built as a concatenation of identical subgraphs (see Figure 5.1). We label the *motifs* which build these graphs as F_n, and for illustration purposes we show in Figure 5.2 the first four of them.

For a fixed n, we can then 'concatenate' motifs (in a way which will be formally defined later) and the graph resulting of concatenating k motifs is denoted by F_n^k (so that $F_n^1 = F_n$ and $\lim_{k\to\infty} F_n^k = F_n^\infty$). Whereas in [84] a Feigenbaum graph was defined for a bi-infinite trajectory ($k \to \infty$), one can however extend this definition to *finite* graphs by fixing a finite k. Accordingly, the elements in the bi-parametric set $\{F_n^k\}_{n\geq 0, k\geq 0}$ (where $F_n^1 := F_n$) provides a useful enumeration of finite Feigenbaum graphs. For completeness, we define F_n^0 to be the empty graph of one node. With a little abuse of language, in what follows we will indistinctively refer to F_n^k and F_n^∞ as Feigenbaum graphs.

Remark. Given an integer n, both F_n^k and F_n^∞ are unique $\forall \mu \in I_n$: for the range of values of μ for which the map is periodic and the associated time series has the same period, the resulting Feigenbaum graph is unique, i.e., it is not dependent on the map's initial condition. This observation, as we shall see, does not hold for the range of values of μ that correspond to chaotic behaviour. Furthermore, the hierarchy of Feigenbaum graphs is universal for all unimodal maps undergoing a Feigenbaum scenario. In particular, this

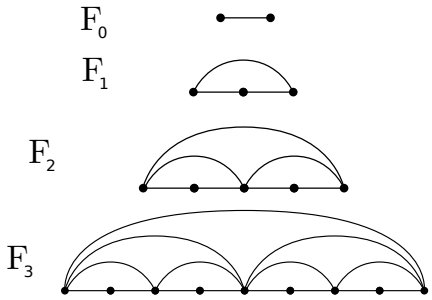


Figure 5.2: Single motifs $F_n^1 \equiv F_n$ of the HVGs associated to the logistic maps with period $T = 2^n$.

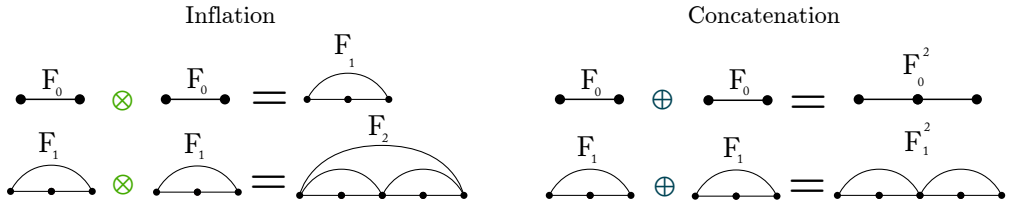


Figure 5.3: A visualisation of the motif inflation ($\otimes$) and motif concatenation ($\oplus$) rules.

means that this hierarchy is not only associated to the logistic map but to any unimodal map. The reason is because Feigenbaum graphs are based in the order of visits to the stable branches and this order is unique for all unimodal maps [84].

Note that we can generate all the elements of the family $\{F_n^k\}_{n,k\geq 0}$ by combining them using two graph-theoretical operations which we now define:

Definition 5.2.2 (Motif inflation $\otimes$)**.** Consider two undirected graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, where V_i are the vertex sets ($|V_i| = N_i$) and E_i are the edge sets, where V_i are totally ordered. We label the vertex set of G_1 by $V_1 = (1, 2, \ldots, N_1)$ and similarly for G_2 we have $V_2 = (1', 2', \ldots, N_2')$. Then $G_1 \otimes G_2$ is a graph which fulfils the following conditions:

1. $G_1 \otimes G_2$ is a graph with $N_1 + N_2 - 1$ vertices,
2. whose vertex set $V'' = (1, 2, \ldots, N_1 - 1, 1'', 2', 3', \ldots, N_2')$,
3. where vertex $1''$ is a block vertex that merges the vertices N_1 and $1'$ (from V_1 and V_2 respectively), and inherits all the edges that were incident to both of them.

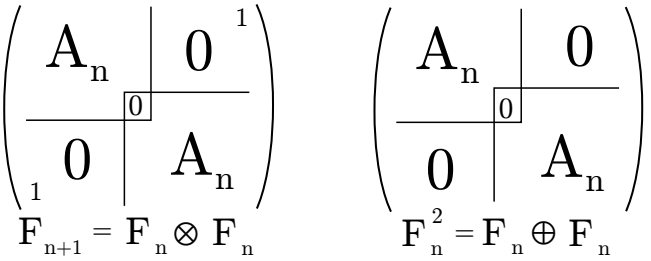


Figure 5.4: The adjacency matrices of $F_{n+1} = F_n \otimes F_n$ and $F_n^2 = F_n \oplus F_n$ expressed in terms of $\mathbf{A}_n$. These composition rules are labelled as a graph inflation and concatenation respectively.

4. The vertices 1 and N_2' share an edge in $G \otimes G$.

5. The remaining edge set is formed by all edges between vertices $\{2, 3, \ldots, N_1 - 1\}$ inherited from G_1 and between the vertices $\{2', 3', \ldots, [N_2 - 1]'\}$ inherited from G_2.

For illustration, a visualisation of the inflation operation is shown in the left panel of Figure 5.3.

It is easy to see that $F_{n+1} = F_n \otimes F_n$. Let us define $\mathbf{A}_n$ as the adjacency matrix of F_n (defining the adjacency matrix $\mathbf{A} = \{a_{ij}\}$ to be a binary matrix which assigns $a_{ij} = 1$ if i and j are two nodes linked by an edge, and zero otherwise). The adjacency matrix $\mathbf{A}_{n+1}$ of F_{n+1} can be expressed in terms of the adjacency matrix $\mathbf{A}_n$ of F_n as illustrated in Figure 5.4. Therefore, starting from $k = 1$, the operation $\otimes$ iteratively generates all the elements of the set $\{F_n^1, n > 0\}$. This means that $(F_n, \otimes)$ is a unary system if we interpret $\otimes$ as a unary operation $\otimes : \{F_n\} \to \{F_n\}$. Notice however that this set is not closed under $\otimes$, as for $n_1 \neq n_2$, $\nexists\, n_3 > 0$ such that $F_{n_1} \otimes F_{n_2} = F_{n_3}$. The graphs formed by combining together F_{n_1} and F_{n_2} with $n_1 \neq n_2$ are not Feigenbaum graphs, but are still HVGs, hence the set of all HVGs is closed under this operation.

Definition 5.2.3 (Motif concatenation $\oplus$)**.** Consider two undirected graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, where V_i are the vertex sets ($|V_i| = N_i$) and E_i are the edge sets, where V_i are totally ordered. We label the vertex set of G_1 by $V_1 = (1, 2, \ldots, N_1)$ and similarly for G_2 we have $V_2 = (1', 2', \ldots, N_2')$. Then $G_1 \oplus G_2$ is a graph which fulfils

the following conditions:

1. $G_1 \oplus G_2$ is a graph with $N_1 + N_2 - 1$ vertices,
2. whose vertex set $V'' = (1, 2, \ldots, N_1 - 1, 1'', 2', 3', \ldots, n_2')$,
3. where vertex $1''$ is a block vertex that merges the N_1 and $1'$ (from V_1 and V_2 respectively), and inherits all the edges that were incident to both of them,
4. The vertices 1 and N_2 **do not** share an edge in $G_1 \oplus G_2$,
5. The remaining edge set is formed by all edges between vertices $\{2, 3, \ldots, N_1 - 1\}$ inherited from G_1 and between the vertices $\{2', 3', \ldots, [N_2 - 1]'\}$ inherited from G_2.

A visualisation of this rule, both graphically and algebraically, is shown in the right panels of Figures 5.3 and 5.4 respectively. We also define $\mathbf{A}_n^k$ to be the adjacency matrix of F_n^k. To avoid confusion, we also state that, using parentheses, $(\mathbf{A}_n^k)^p$ is the pth power of the corresponding adjacency matrix. One can easily see that, locally, the inflation rule on two graphs $G_1 \otimes G_2$ is equivalent to the concatenation one $G_1 \oplus G_2$ if we add an extra edge between the first and last vertex (see Figure 5.3 for illustration). Now, for any given n, one has $F_n^k \oplus F_n^1 = F_n^{k+1}$. It is also easy to prove that $\forall n \geq 0,\ k_1, k_2 \geq 0, F_n^{k_1} \oplus F_n^{k_2} = F_n^{k_1+k_2}$. Therefore, for a fixed $n > 0$, the operation $\oplus$ generates all the elements of the set $\{F_n^k, k > 0\}$. It is also easy to prove that, for a fixed $n > 0$, $(F_n^k, \oplus)$ is a commutative monoid with the identify element being the empty graph of one node F_n^0, hence $(F_n^k, \oplus)$ is isomorphic to $(\mathbb{N}, +)$.

Remark. Note that the set $\{F_n^k\}|_{n,k}$ can also be created with the aid of simplicial complexes. Given an arbitrary F_n $(n > 1)$, we can create F_{n+1} by gluing a triangle (i.e., a 2-simplex) to the edges attached to each node with degree 2. In Figure 5.5 we show F_3 in the standard way, along with the corresponding simplicial complex representation and equivalent nodes. In the same Figure we also depict F_4, with the eight 2-simplices glued to the edges attached to each node with degree 2 in F_3. Incidentally, note that

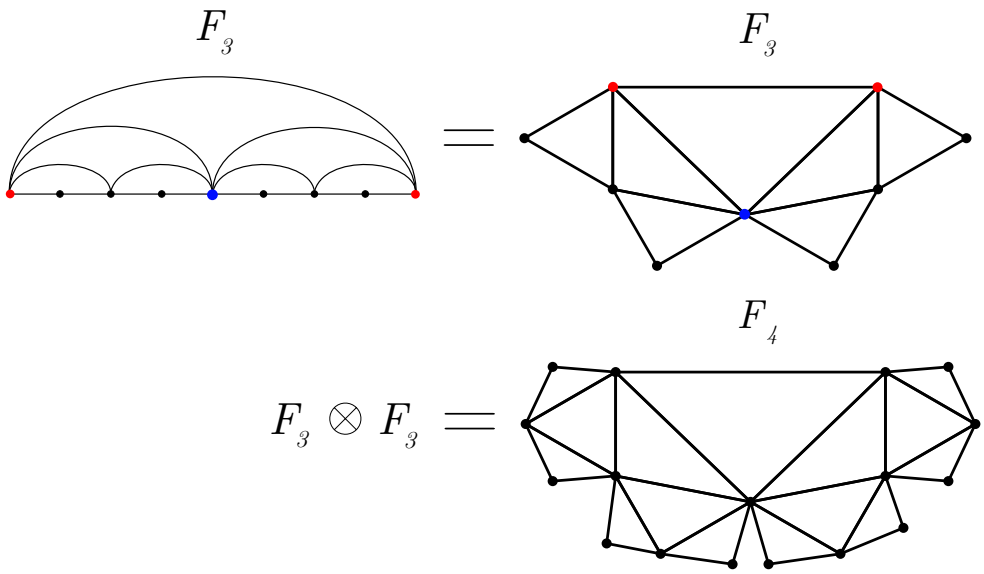


Figure 5.5: A visualisation of the standard layout of F_3 (top left) with its simplicial complex layout (top right), along with equivalent nodes in red and blue. On the bottom we plot F_4, and we can see the 8 2-simplices glued to the edges attached to each node with degree 2.

this visualisation reveals some similarities between the family $\{F_n\}$ and the Geometric Assortative Growth Model [86].

The set $\{F_n^k\}|_n$ (fixed n) is finitely generated by F_n^1 under $\oplus$, while the generating set of $\{F_n\}$ under $\otimes$ is $\{F_1\}$. Now, consider the larger set $\{F_n^k, n, k \in \mathbb{N}\}$ where n and k are now free parameters. This set contains the two-parameter (n, k) family of Feigenbaum graphs. This set is again finitely generated by F_1 using the operations $\oplus$ and $\otimes$. Exploration of the algebraic properties of $\{F_n^k, n, k \in \mathbb{N}\}$ is an interesting topic for future research. However, here we are interested in the spectral properties of F_n^k. Some very basic observations, which will be helpful later in the task of bounding eigenvalues, are summarised in the following proposition.

Proposition 5.2.4. Consider the set of graphs F_n^k, for $n, k \in \mathbb{N}$, and let V_n^k and E_n^k be the size of the vertex and edge set respectively, with $V_n := V_n^1$, $E_n = E_n^1$. Then the following holds:

1. F_n is a graph with $V_n = 2^n + 1$ vertices and $E_n = 2^{n+1} - 1$ edges.

2. F_n^k is a graph with $V_n^k = 2^n k + 1$ vertices and $E_n^k = k(2^{n+1} - 1)$ edges.

Proof. The proof trivially follows from the definitions of $\otimes$ and $\oplus$. □

The spectral properties of F_n^k will be addressed in Section 5.3. For readability, we will split this initial study in two natural directions: in Section 5.3.2 we set $k = 1$ and consider the spectral properties of F_n (i.e., for $k = 1$, as n increases), whereas in Section 5.3.4 we set n fixed and consider the spectral properties of F_n^k as k increases, i.e., the finite size truncations of infinite Feigenbaum graphs. Finally in Section 5.3.6 we will explore the spectrum of F_n^k when n and k are finite and both vary.

5.2.2 The large n and k limits

The variables n and k have clear, different meanings: n is related to the period T of the logistic map's trajectories via $T = 2^n$ (physically speaking, n is related to μ). In particular, the period-doubling bifurcation cascade that the logistic map experiences relates to successive increases of n, where the onset of chaos ($\mu = \mu_\infty$) is only reached in the limit $n \to \infty$. On the other hand, k is a parameter that only describes the length of the trajectory (and therefore properties of a trajectory, for example its periodicity, will only be revealed when k is large, or in the limit $k \to \infty$), in particular k is the number of concatenated motifs and is related to the size N of the trajectory (the length of the time series) via $N = V_n^k = 2^n k + 1$. Note that a priori we have two possible ways to take the limits of large n and k. On the one hand, we can fix n and let $k \to \infty$. This mimics the situation where we have an infinitely long trajectory of finite period $T = 2^n$. In this limit, F_n^k is by construction a locally finite infinite graph, i.e., the number of vertices is infinite but each vertex has a finite number of edges.

On the other hand, we can also fix k (e.g., $k = 1$) and take $n \to \infty$. This mimics the situation where only a single 'period' is extracted from the series, however as this period is $T = 2^n$, in the limit the time series is infinitely long, obtaining an infinite graph. However, in this limit the graph is *not* locally finite: as we will show later in

Proposition 5.3.3 the degree of the central vertex of F_n increases linearly with n, so there are at least k vertices in F_n^k whose degree increases (without bound) with n. On the other hand, this is still a countable infinite graph.

Therefore, taking the limits $k \to \infty$ and $n \to \infty$ yield different types of infinite graphs: a locally finite infinite graph on one hand and a countable infinite graph on the other. In particular, the fact that the limit $n \to \infty$ yields infinite graphs which are not locally finite has important consequences for the spectral properties of these graphs. Recall that for finite graphs, the spectrum of a graph is simply the set of all eigenvalues of the respective adjacency matrix $\mathbf{A}$. However if the graph is infinite, the spectrum of $\mathbf{A}$ depends on the choice of the space on which $\mathbf{A}$ acts as a linear operator (typically one considers the Hilbert space $\ell^2(V)$, where V is the set of vertices). It is well known that if the infinite graph is locally finite, then $\mathbf{A}$ acts on $\ell^2(V)$ as a self-adjoint operator and its norm is smaller or equal to $d_{\max}$, the largest degree of the graph [87]. However, if the property of local finiteness is relaxed then this operator no longer bounded and one needs to compactify it before attempting to use any spectral theory.

In summary, the limit $\lim_{n,k\to\infty} F_n^k$ (which is the one we should take to explore the onset of chaos $\mu = \mu_\infty$) is non-trivial. For this reason, we leave these as interesting open problems, and from now on we will assume that both n and k are arbitrary large but finite.

5.2.3 Feigenbaum graphs with $\mu > \mu_\infty$: Chaotic Feigenbaum graph ensembles

In the range $\mu > \mu_\infty$, the trajectories of the logistic map are typically chaotic (except for the windows of periodicity, which are essentially subintervals where the period-doubling cascade is self-similarly reproduced albeit with an initial period larger than one). The first observation is that in the chaotic regime the graphs can no longer easily be enumerated. In fact, for a given μ in the chaotic range the Feigenbaum graph is no longer unique: each different condition will typically generate a different chaotic trajectory and

therefore a different Feigenbaum graph. Hence each value of μ spans a different *ensemble* of Feigenbaum graphs, generated by sampling different initial conditions in the map. As discussed in Section 5.2.1, this is at odds with the case $\mu < \mu_\infty$, where for any particular μ all realisations in an ensemble associated to μ yielded the same Feigenbaum graph, therefore the ensemble was fully degenerate in that case.

Of course as the length of the time series approaches infinity, the statistical properties of two different chaotic trajectories extracted at the same value of μ are asymptotically identical, so we expect some kind of statistical equivalence in the resulting Feigenbaum graphs. For instance, for $\mu = 4$ (fully developed chaos) one can compute the degree distribution of the (ensemble of) Feigenbaum graphs. This is a statistical quantity which can be solved analytically by using a diagrammatic technique [36], and has been shown to be a valid limit for single realisations. However, here we are interested in studying the spectral properties of Feigenbaum graphs, so we need to address whether these properties are sufficiently 'robust', i.e., we should check whether these properties do not change much between realisations. This naturally leads to the concept of *self-averaging quantities*, which will be investigated in Section 5.4.1. Then, in Section 5.4.2 we will try to relate the properties of the time series to spectral properties of the graphs.

5.3 The case $\mu < \mu_\infty$: Spectral properties in the period-doubling cascade

Here we explore the spectral properties of $\{F_n^k\}$. In particular, we will focus on the maximal eigenvalue of the adjacency matrix of F_n^k, although other properties will also be considered, such as the full spectrum, the determinant, and the tree number. For convenience, we split this section in three main blocks: the first explores the dependence of n by focusing on the properties of $\{F_n\}_{n\geq 0}$. The second focuses on the dependence on k by exploring properties of $\{F_n^k\}_{k\geq 0}$ where n is fixed. Finally we explore $\{F_n^k\}_{n\geq 0,k\geq 0}$, where both n and k can vary.

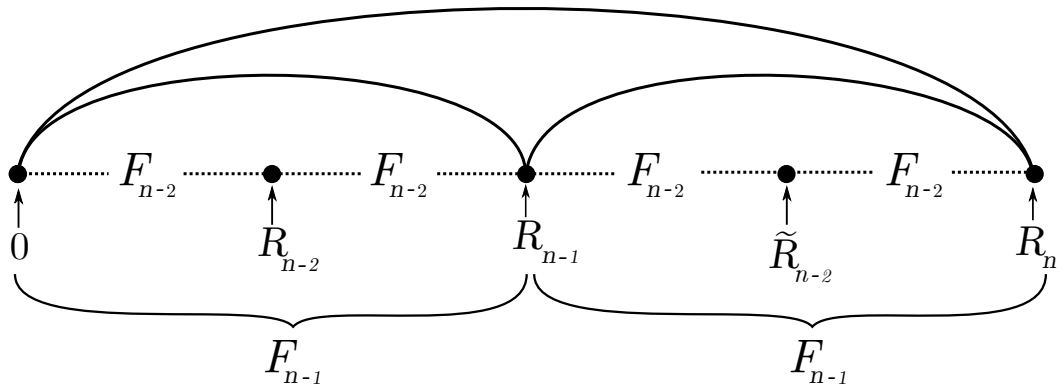


Figure 5.6: A diagram showing the notation used in Theorem 5.3.1. The extremal points are labelled R_n, with the extremal points of the previous Feigenbaum graphs (which F_n has embedded, and are also labelled in the figure) labelled as R_{n-1}, etc. The extremal point of the middle of the right side is labelled as $\tilde{R}_{n-2}$ (although going deeper than this would cause a notation issue, but this is sufficient to complete the proof).

5.3.1 A first view on the full spectrum of F_n

Here we fix $k = 1$ and consider the set $\{F_n\}_{n\geq 0}$, and we start by exploring the full spectrum the adjacency matrices $\{\mathbf{A}_n\}_{n\geq 0}$ (i.e., the set of $2^n + 1$ eigenvalues) associated to F_n. The first quantity worth exploring is the number of distinct eigenvalues of $\mathbf{A}_n$, labelled $q(\mathbf{A}_n)$. To bound this, it is useful to resort to the *diameter* D_n of F_n, defined as $D_n = \max_{i,j}\{\delta_{ij}\}$, where δ_{ij} is the (shortest path) distance between node i and node j. A well known result is $q(\mathbf{A}_n) \geq D_n + 1$. The following theorem provides the diameter D_n:

Theorem 5.3.1. *(Diameter of F_n) The diameter of F_n is $D_n = n$.*

Proof. The proof requires a Lemma. Let us consider F_n, whose nodes are labelled $\{0, 1, 2, \ldots, 2^n\}$, and denote by $R_n = 2^n$ the rightmost node of F_n. In the following, we call *extremal points* of F_n the nodes 0, R_{n-1}, and R_n. Notice that R_{n-1} is the middle-point of the Hamiltonian path of F_n^1 that starts at node 0 and proceeds by increasing node labels. Notice as well that since $F_n = F_{n-1} \otimes F_{n-1}$, F_{n-1} has again three extremal points, which are labelled (with respect to F_n) either $\{0, R_{n-2}, R_{n-1}\}$ or $\{R_{n-1}, \tilde{R}_{n-2}, R_n\}$. Refer to Figure 5.6 for a diagram showing this notation. This proof

is less complex than it seems, and is far more easy to read if one follows along with Figure 5.6. We denote by δ_{ij} the distance between node i and node j, i.e., the length of the shortest path from i to j. We can prove the following Lemma

Lemma 5.3.2. *(Distance to the closest extremal point.) Consider the graph F_n and the minimal distance between a generic node i and the closest extremal point E*

$$\delta^n_{i,E} = \min\{\delta_{i,0}; \delta_{i,R_{n-1}}; \delta_{i,R_n}\}.$$

Then for $n \geq 1$ we have:

$$\delta_{i,E} \leq \left\lfloor \frac{n}{2} \right\rfloor, \forall i \in [0, R_n].$$

Proof. We will prove this by strong induction on n. When $n = 1$, F_n is a triangle whose diameter is $D_1 = 1$ and thus all three nodes are extrema, i.e., $\delta_{i,E} = 0 = \lfloor \frac{1}{2} \rfloor$. Now let us assume that the Lemma is valid up to $n-1$ and let's prove it for n. Without loss of generality, we assume that node i belongs to the left-most half of F_n, such as $i \in [0, R_{n-1}]$. In this case we have that $\delta^n_{i,E} = \min\{\delta_{i,0}, \delta_{i,R_{n-1}}\}$, since R_n will be at least one hop farther away from i than either 0 or R_{n-1}. Let us consider first the case where $i \in [0, R_{n-2}]$. In this case i is closer to 0 than to R_{n-1} (or at most, at the same distance from either of the two), hence $\delta^n_{i,E} = \min\{\delta_{i,0}, \delta_{i,R_{n-1}}\} = \delta_{i,0} \leq 1 + \delta^{n-2}_{i,E} \leq 1 + \lfloor \frac{n-2}{2} \rfloor$. The first inequality is due to the fact that node 0 is an extremal point, and the distance from i to 0 will be either equal to $\delta^{n-2}_{i,E}$ or to $1 + \delta^{n-2}_{i,E}$. The second inequality is due to the induction assumption that (Lemma 5.3.2) is valid up to $n-1$. If n is even, we have: $\delta^n_{i,E} \leq 1 + \frac{n-2}{2} = \frac{n}{2} = \lfloor \frac{n}{2} \rfloor$. If n is odd instead, we have: $\delta^n_{i,E} \leq 1 + \frac{n-3}{2} = \frac{n-1}{2} = \lfloor \frac{n}{2} \rfloor$. The case where $i \in [R_{n-2}, R_{n-1}]$ is similar, since we can relabel each node in $[R_{n-2}, R_{n-1}]$ according to the function $\phi(i) = i - R_{n-2}$, and repeat the same reasoning. In conclusion, $\delta^n_{i,E} \leq \lfloor \frac{n}{2} \rfloor$ for all $i \in [0, R_{n-1}]$. But since the graph is symmetric around R_{n-1}, we have $\delta^n_{i,E} \leq \lfloor \frac{n}{2} \rfloor$ for all i in $[0, R_n]$. □

We can now finish the proof of Theorem 5.3.1. Let us consider two generic nodes i and

j in F_n. First consider the case where $i \in [0, R_{n-1}]$ and $j \in [R_{n-1}, R_n]$. We have two possibilities for $\delta_{i,E}$ (either $\delta_{i,E} = \delta_{i,0}$ or $\delta_{i,E} = \delta_{i,R_{n-1}}$) and two possibilities for $\delta_{j,E}$ (either $\delta_{j,E} = \delta_{j,R_n}$ or $\delta_{j,E} = \delta_{j,R_{n-1}}$). So we have that

$$\delta_{ij} = \min \begin{cases} \delta_{i,0} + 1 + \delta_{j,R_n}, \\ \delta_{i,0} + 1 + \delta_{j,R_{n-1}}, \\ \delta_{i,R_{n-1}} + 1 + \delta_{j,R_n}, \\ \delta_{i,R_{n-1}} + \delta_{j,R_{n-1}}. \end{cases}$$

This yields

$$\begin{aligned} \delta_{ij} &= \min\{\delta_{i,E} + 1 + \delta_{j,E}, \delta_{i,E} + \delta_{j,E}\} \\ &= \delta_{i,E} + \delta_{j,E} \\ &\leq 2 \left\lfloor \frac{n}{2} \right\rfloor \leq n. \end{aligned}$$

where we have used (Lemma 5.3.2). Conversely, if we have that $i, j \in [0, R_{n-1}]$ (or equivalently, both $i, j \in [R_{n-1}, R_n]$) then:

$$\delta_{ij} = \min \begin{cases} \delta_{i,R_{n-1}} + 1 + \delta_{j,0}, \\ \delta_{i,0} + 1 + \delta_{j,R_{n-1}}, \\ \delta_{i,0} + \delta_{j,0}, \\ \delta_{i,R_{n-1}} + \delta_{j,R_{n-1}}. \end{cases}$$

With a similar argument as above, we get

$$\begin{aligned} \delta_{ij} &= \min\{\delta_{i,E} + d_{j,E}, 1 + \delta_{i,E} + \delta_{j,E}\} \\ &= \delta_{i,E} + \delta_{j,E} \\ &\leq 2 \left\lfloor \frac{n}{2} \right\rfloor \leq n, \end{aligned}$$

i.e., $\delta_{ij} \leq n$. Now, for an arbitrary n, we can always find a pair of nodes l, m in F_n which saturates the inequality, with $\delta_{l,m} = n$. For example, in F_2 we can set l equal to

node 1, and m equal to node 3, and for F_3 we have $(l, m) = (1, 5)$ and for F_4 we have $(l, m) = (2, 12)$. To construct an algorithm that provides l and m in the general case, we start with F_3, pictured in the top left panel of Figure 5.7, along with nodes l and m, with a shortest path (which is not unique) between them coloured in red. We move to the top right panel, where we have F_4 overlaid with the additional edges highlighted with dotted lines. We can move m to m', and the length of the shortest path between l and m' is increased by 1. This is because the new 2-simplex which we moved m in to is *not* glued to an edge which is a member of a shortest path between l and m. We can repeat this process when moving from F_4 (bottom left panel) to F_5 (bottom right panel), however instead of moving m', we move l in a similar fashion. This is because the new simplices that are glued to the edges of m' are a member of a shortest path between l and m', hence if we were to again move m' to the new 2-simplex joined to it, we would not increase the length of the shortest path. But if we move l to l' we again increase the length of the shortest path between our nodes by 1. Repeating this process (by induction, using F_3 as our base case), alternating the movement of l and m, we can find a shortest path between any two nodes of F_n, with length n.

Hence we can always find a l and m to give δ_{ij}, and combining this with the bound $\delta_{ij} \leq n$ we have that $\max_{i,j}\{\delta_{i,j}\} = n$, which concludes the proof of Theorem 5.3.1. □

According to the theorem above, we conclude $q(\mathbf{A}_n) \geq n+1$. To evaluate how tight this bound is, we first order the eigenvalues of $\mathbf{A}_n$ in ascending order as

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_r \leq \cdots \leq \lambda_{2^n+1},$$

where r is just a rank label. In Figure 5.8 we plot the entire (point) spectrum of F_n for $n \leq 10$. We can make several observations:

(i) First, the bound on $q(\mathbf{A}_n)$ provided above does not seem to be tight, when comparing to the numerical evidence. On the contrary, the numerical evidence suggests instead that $q(\mathbf{A}_n) = V_n = 2^n + 1$, i.e., all eigenvalues seem to be *distinct*, something that we leave as a conjecture.

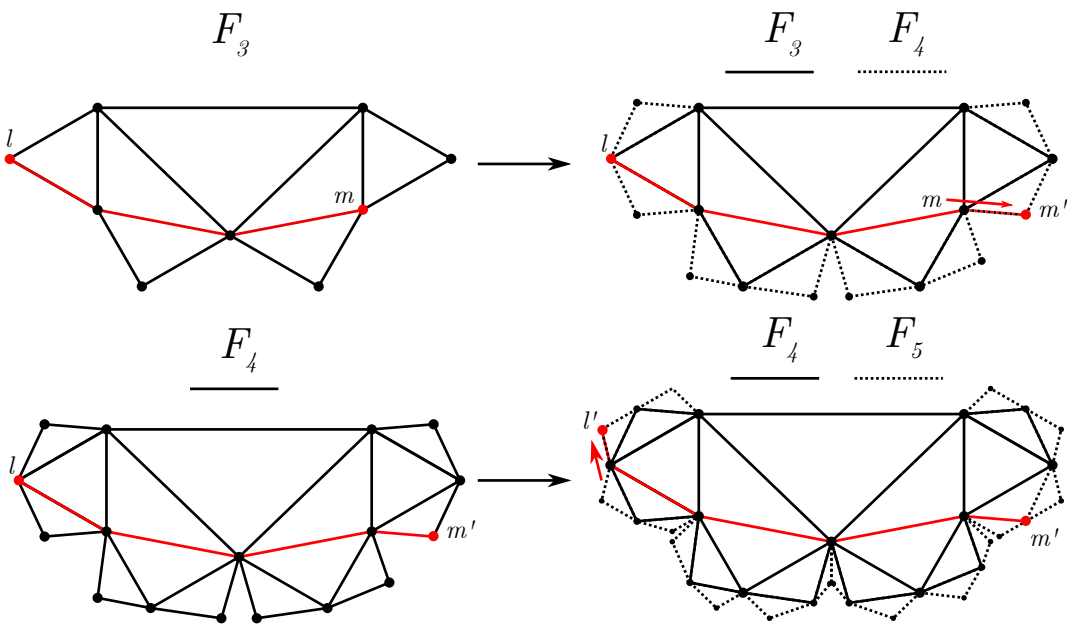


Figure 5.7: A visualisation of the algorithm used to create a shortest path F_n, with length n. In the top left panel we have F_3 with a shortest path between nodes l and m highlighted in red. In the top right panel we show how a new path can be created, in F_4, with length 4, by moving m. In the bottom panels we show how we can move l to create shortest path in F_5 with length 5. The algorithm is described in the text.

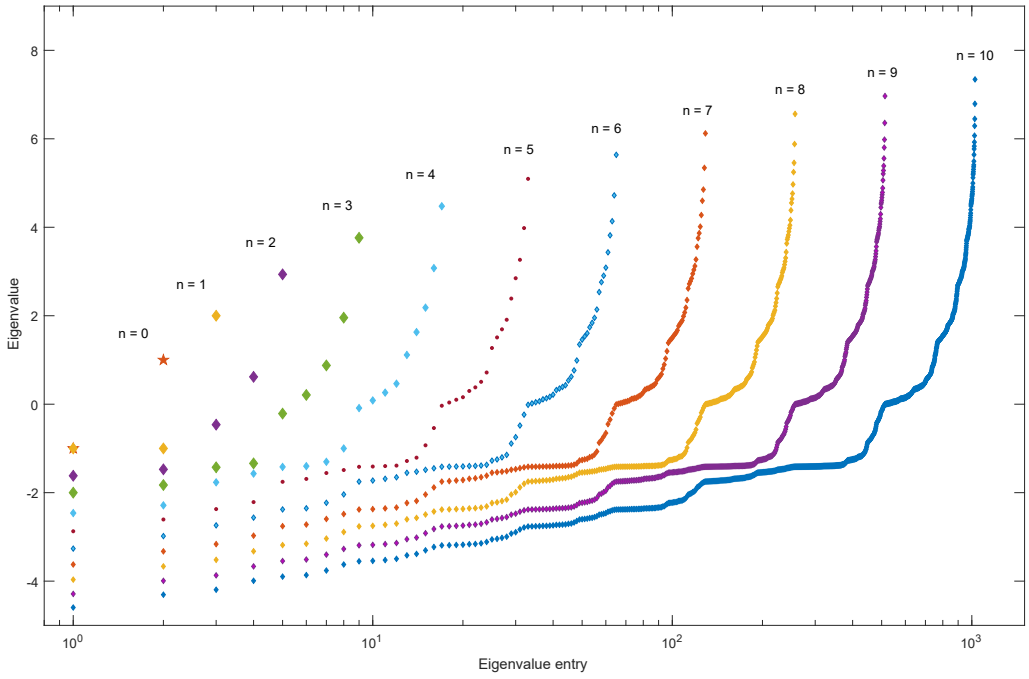


Figure 5.8: A semilog plot of the spectrum (sorted in ascending order) of F_n for $0 \leq n \leq 10$. The plot suggests that the spectrum seems to be converging to a particular curve and that all the eigenvalues are distinct.

(ii) Second, the spectrum is not symmetric and in particular the largest ($\lambda_{\max}$) and smallest ($\lambda_{\min}$) eigenvalues are different in modulus (thanks to the Perron-Frobenius theorem for primitive matrices, as discussed in Section 5.3.2.1).

(iii) Third, the ordered spectrum seems to be converging to a peculiar, limiting shape as n increases. In order to investigate this apparent convergence, in Figure 5.9 we report the rescaled spectra for $n \in [2, 13]$. We considered the spectrum of F_{13} as a reference (hence the range on the x axis), which has $2^{13} + 1 = 8193$ distinct eigenvalues. In order to appropriately compare the spectra of each graph (which have different number of eigenvalues), we have rescaled the rank label of each eigenvalue using the function:

$$R(r_n) = (r_n - 1)2^{m-n} + 1 \tag{5.2}$$

where r_n is the rank label of an eigenvalue of F_n (i.e., $r_n \in [1, 2^n + 1]$), and m is the order of the spectrum taken as a reference (in this case, $m = 13$ for F_{13}). Notice that, as a limiting case, the transformation in Equation (5.2) maps the rank label of the smallest eigenvalue λ_1 of F_n into 1, and the rank label of the largest eigenvalue λ_{2^n+1} of F_n into $2^m + 1$ (that for $m = 13$ is equal to 8193). Indeed, the gap between the rescaled rank labels of two subsequent eigenvalues of F_n is equal to $\left\lfloor \frac{2^m}{2^n+1} \right\rfloor - 1$. This transformation ensures that the record statistics of all the spectra will be mapped to the same record statistics of F_m. For instance, the median eigenvalue of the spectrum of F_n, which has rank label $2^{n-1} + 1$, will be mapped into the rank label $2^{m-1} + 1$, that for $m = 13$ corresponds to the rank label 4097. Similarly, the eigenvalue corresponding to the upper-quartile of the spectrum of F_n will always be mapped to the rank label 6145, and so on. It is evident from Figure 5.9 that thanks to this rescaling all the spectra seem to nicely collapse towards a mysterious limiting curve. This result is somehow related to the fact that F_{n-1} is always a proper subgraph of F_n, so that the eigenvalues of F_{n-1} are interlaced with those of F_n as a consequence of Cauchy's interlacing theorem [88]. However, the shape of the limiting curve has defied all our attempts at an analytic description so far.

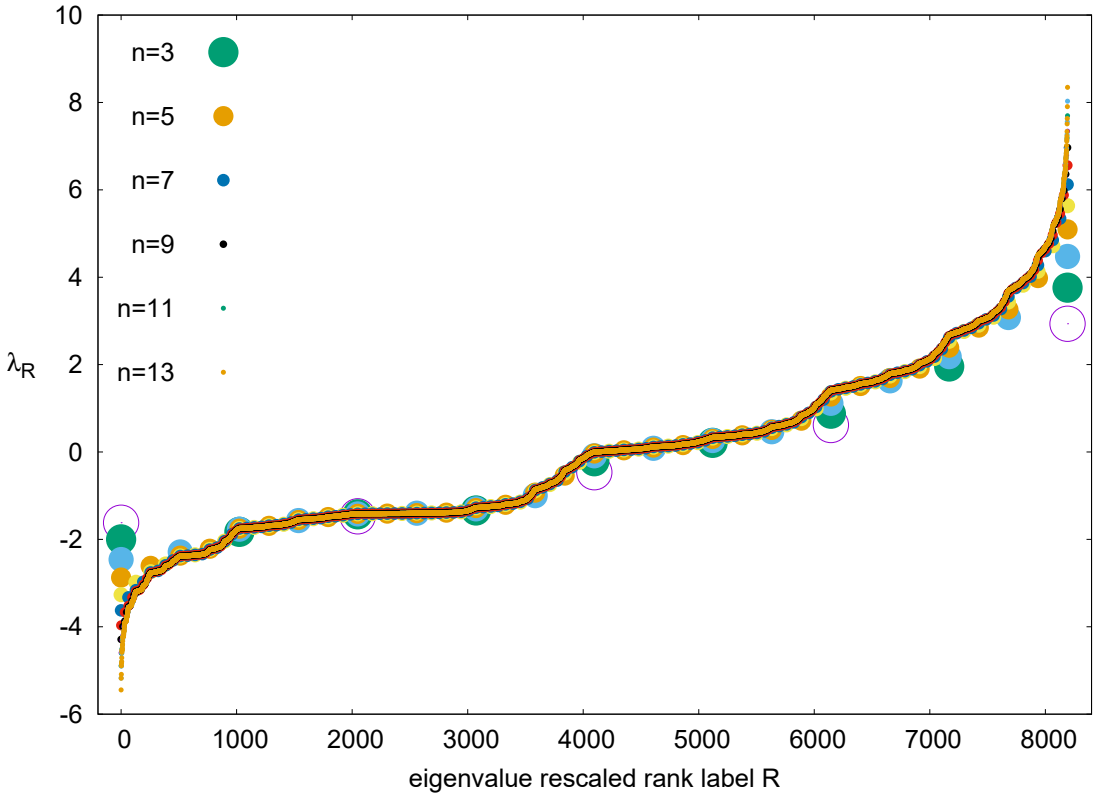


Figure 5.9: The collapse of all the spectra displayed in Figure 5.8 reveals that, after an appropriate rescaling, the spectrum of F_n converges to a limit function. The point size is inversely proportional to n, so that larger dots are used for the eigenvalues of F_n corresponding to smaller values of n, and vice versa.

5.3.2 Largest eigenvalue $\lambda_{\max}$ for F_n

Here we continue to focus on the case $k = 1$, and turn our attention to the largest eigenvalue of $\mathbf{A}_n$. As $\mathbf{A}_n$ is irreducible (the graph F_n is undirected and connected), according to the Perron-Frobenius for non-negative irreducible matrices, the largest eigenvalue λ_{2^n+1} has multiplicity 1, we then define $\lambda_{\max}(F_n) = \lambda_{2^n+1}(\mathbf{A}_n)$ (and similarly we define $\lambda_{\min}(F_n) = \lambda_1(\mathbf{A}_n)$).

Using the `eigs` function in MATLAB it is possible to efficiently calculate the largest eigenvalue of sparse matrices, even if the matrices are large. In Figure 5.10 we plot, in a log-log scale, $\lambda_{\max}(F_n)$ for $1 \leq n \leq 26$. The data fit very well to the power law dependence $\lambda_{\max} \sim n^{\alpha}$ with $\alpha \approx 0.5$.

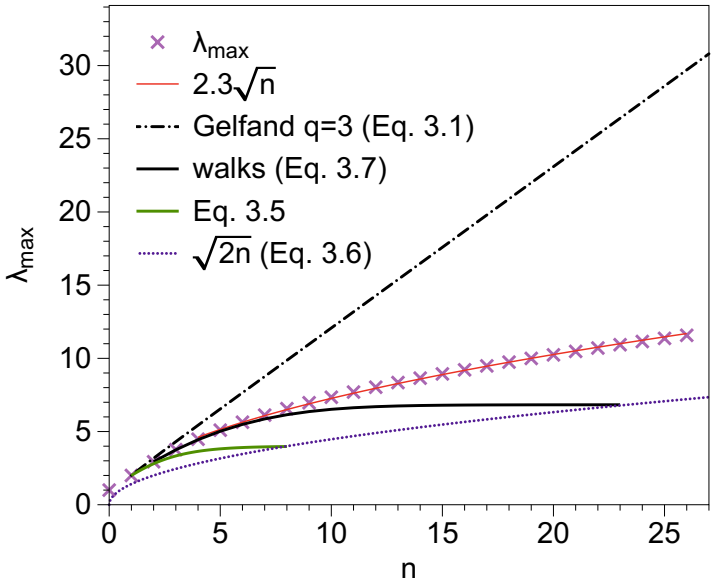


Figure 5.10: Crosses denote numerical computation $\lambda_{\max}(F_n)$, for $n = 1, \ldots, 26$ (note that F_{26} has over 67 million nodes). The red solid line provides a power law fit $\lambda_{\max} \sim n^{1/2}$. The rest of the lines are different analytical upper and lower bounds (see the text).

In this section our aim is to explain this scaling by finding adequate bounds.

5.3.2.1 Gelfand's formula

Gelfand's formula provides a bound for the spectral radius of an adjacency matrix A:

$$\rho(A) = \lim_{q \to \infty} ||A^q||^{1/q},$$

where $||\cdot||$ is any matrix norm. In particular, for any finite $q \in \mathbb{Z}^+$ we have that $\rho(A) \leq ||A^q||^{1/q}$.

It is easy to prove that $\rho(\mathbf{A}_n) = \lambda_{\max}(F_n)$. In fact, $\mathbf{A}_n$ is non-negative and irreducible, since the graph F_n is non-empty, undirected and connected, and is also aperiodic, since each node of the graph belongs to at least one triangle. Consequently, $\mathbf{A}_n$ is primitive. The Perron-Frobenius theorem for non-negative primitive matrices guarantees that the largest eigenvalue is real, simple, and equal to the spectral radius $\rho(A)$. Therefore we can write

$$\rho(\mathbf{A}_n) = \lambda_{\max} \leq ||A^q||^{1/q}.$$

For simplicity, we choose $||\cdot||_\infty$, defined as $||A||_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$. We have that $||(\mathbf{A}_n)^1|| = 2n$, this is because the node with the largest amount of 1-walks is the node with the largest degree; this is the central node and has degree $2n$ (see Proposition 5.3.3 below). For $q = 2$ we have $\|(\mathbf{A}_n)^2\| = 2n^2 + 2$, a result which we prove in Section 5.3.5. For $q = 3$ we calculate $\|(\mathbf{A}_n)^3\|$ for several values of n and numerically find that they exactly fit a cubic equation:

$$\|(\mathbf{A}_n)^3\| = \frac{4}{3}n(n+1)(n+2)$$

We did not find a closed formula for $4 \leq q \leq 10$ (although one may very well exist). Taking respectively the 1st, 2nd and 3rd roots of these three formulas, we find that for $q < 4$ the approximant to the spectral radius is essentially linear on n, providing our first estimated upper bound for $\lambda_{\max}$. Because we did not find a closed expression for $4 \leq q \leq 10$, we take $q = 3$ as our 'Gelfand's estimate' and we have a conjecture:

$$\lambda_{\max} \leq \left[\frac{4}{3}n(n+1)(n+2)\right]^{1/3}. \tag{5.3}$$

5.3.2.2 Bounds on largest eigenvalue based on degree.

In order to improve the bound provided by Gelfand's formula, we now turn to the specific bounds for the largest eigenvalue that exist in the literature. Some elementary bounds for the largest eigenvalue of a graph G with maximum degree $d_{\max}$ and average degree $\bar{d}$ [89] are:

$$\max\{\bar{d}, \sqrt{d_{max}}\} \leq \lambda_{\max} \leq d_{\max} \tag{5.4}$$

$$\lambda_{\max} \leq \max\{\sqrt{d_i d_j} : 1 \leq i, j \leq n, v_i v_j \in E\}, \tag{5.5}$$

where E is the edge set. We apply these bounds to F_n. We summarise the bounds in the following proposition:

Proposition 5.3.3. Consider F_n. Then

(a) The largest degree of F_n is found in its central vertex and is $d_{\max}(F_n) = 2n$.

(b) The vertices with second largest degree are the boundary ones (first and last) and each have degree $n+1$.

(c) The average degree is $\bar{d}(F_n) = 4 - 6/(2^n+1)$

(d) $\max\{\sqrt{d_i d_j} : 1 \leq i,j \leq n, v_i v_j \in E\} = \sqrt{2n(n+1)}$

Proof. First, observe that for $n \geq 1$ we have $F_n = F_{n-1} \otimes F_{n-1}$, and in the inflation process the only vertices whose degree increases are the border ones (leftmost and rightmost). Proofs of (a) and (b) are then by induction on n: For $p = 1$ we have that $d_{\max}(F_p) = 2$, found in the central vertex, and similarly for the first and last vertex $d = 2$ as well. Then,

- Assume $d_{\max}(F_p) = 2p$. For $n = p+1$, by construction we have $F_{p+1} = F_p \otimes F_p$, so the only vertices that acquire new edges are at the borders of F_p. In particular, the central vertex in F_{p+1} is the one acquiring more edges, and by construction this vertex is built merging the rightmost and leftmost vertex of F_p, hence the central vertex of F_{p+1} has degree $2p + 2p = 2(p+1)$. This finishes the proof for (a).

- Assume that the border vertices (leftmost and rightmost) in F_p have degree $p+1$. In $F_{p+1} = F_p \otimes F_p$, inflation adds an additional edge between the leftmost vertex in the first copy of F_p and the rightmost vertex in the second copy of F_p, and therefore the degree for these nodes in F_{p+1} is just $p+1+1$, finishing the proof for (b).

Moreover, a proof for (c) directly follows from Proposition 5.2.4 by remarking that $\bar{d}(F_n) = 2E_n/V_n$.

Finally, the vertices with largest degree in F_n are the central vertex, with degree $2n$, and the leftmost and rightmost vertices, each of them having degree $n+1$ as previously proved. By construction the central vertex in F_n is always linked with the leftmost and rightmost vertices, hence the identity (d) holds. □

In summary, we find the following bounds based on the degree for $\lambda_{\max}$:

$$\lambda_{\max} \leq \sqrt{2n(n+1)}, \tag{5.6}$$

$$\bar{d}(F_n) = 4 - 6/(2^n + 1) \leq \lambda_{\max}, \text{ for } n < 9, \tag{5.7}$$

$$\sqrt{2n} \leq \lambda_{\max}, \text{ for } n \geq 9. \tag{5.8}$$

Note that asymptotically the lower bound is already $\sim n^{1/2}$ and is therefore tight, whereas the upper bound $\sqrt{2n(n+1)}$ is still linear and worse than our estimate derived from Gelfand's formula.

5.3.2.3 Bounds on largest eigenvalue based on walks.

There exists a general bound for $\lambda_{\max}$ based on number of walks on the graph up to order 3. Let $a(n)$, $b(n)$, $c(n)$ and $d(n)$ be the total number of 3-walks, 2-walks, 1-walks and 0-walks respectively (observe that $d(n)$ is simply the number of vertices and $c(n)$ is just twice the number of edges). Then a lower bound is [90]:

$$\lambda_{\max} \geq \frac{d \cdot a - b \cdot c + \sqrt{d^2 \cdot a^2 - 6abcd - 3c^2 \cdot b^2 + 4(ac^3 + db^3)}}{2(bd - c^2)}. \tag{5.9}$$

In Section 5.3.5 we provide a proof for $b(n), c(n)$ and $d(n)$ along with an estimation for $a(n)$. According to these, we state that for F_n:

$$\begin{aligned} a(n) &= 160 \cdot 2^n - 4n^3 - 30n^2 - 104n - 158 \\ b(n) &= 24 \cdot 2^n - 2n^2 - 12n - 22 \\ c(n) &= 4 \cdot 2^n - 2 \\ d(n) &= 2^n + 1 \end{aligned} \tag{5.10}$$

Based on the leading terms of a, b, c and d it is clear that the lower bound in Equation (5.9) is asymptotically constant, and is therefore a very loose bound for large n. However the expression is extremely good for small values of n, as shown in Figure 5.10.

Summing up, we have exploited different properties such as spectral radius, degree and

walks, and we have obtained several possible bounds accordingly (see Equations (5.3) to (5.9)). The best upper bound is the Gelfand estimate (Equation (5.3)) which is nonetheless still a loose bound. On the other hand the best lower bound is given by the walks bound (Equation (5.9)) for $n < 24$ and by the degree bound (Equation (5.8)) for $n \geq 24$. The scaling of this latter bound seems to be tight. These bounds have been displayed, along with the numerical estimate of $\lambda_{\max}$, in Figure 5.10.

5.3.3 Other spectral properties of F_n: The Tree Number

The tree number of a graph G is the total number of spanning trees, and we will denote it by $\kappa(G)$. To calculate $\kappa(F_n)$ we make use of Kirchhoff's theorem, or the matrix tree theorem:

Theorem 5.3.4 (Kirchhoff's theorem (The Matrix Tree Theorem); [91])**.** *For a given connected graph G with n labelled vertices, let $\mu_1, \mu_2, \ldots, \mu_{m-1}$ be the non-zero eigenvalues of its Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the degree matrix (a diagonal matrix with vertex degrees on the diagonals). Then the number of spanning trees of G is given by*

$$\kappa(G) = \frac{\prod_{i=1}^{m} \mu_i}{m}$$

We have numerically computed the number of spanning trees in F_n, $\kappa(n)$ for $n = 1, \ldots, 4$, the results are shown in Table 5. Interestingly, OEIS states that this sequence (A144621) corresponds to the number of oriented spanning forests of the regular ternary tree with depth n that are rooted at the boundary (i.e., all oriented paths end either at a leaf or at the root), which is given by the recurrence

$$\alpha_{n+2} = \alpha_{n+1}(3\alpha_{n+1} - 2\alpha_n^2) \quad \text{where} \quad \alpha_0 = 0,\ \alpha_1 = 1. \tag{5.11}$$

Hence we conjecture that $\kappa(n) = \alpha_{n+1}$, something which we can easily prove.

Proposition 5.3.5. The number of spanning trees of F_n is $\kappa(n) = 3\kappa(n-1)^2 - 2\kappa(n -$

$1)\kappa(n-2)^2$

Proof. Let's split the contribution of $\kappa(n)$ into two terms. The first part comes from wrongly assuming $F_n = F_{n-1} \oplus F_{n-1}$ instead of the correct $F_n = F_{n-1} \otimes F_{n-1}$ (i.e, we temporarily remove the external bounding edge connecting the first and last node), and the second term comes from explicitly using in the spanning tree this bounding edge. It is easy to see that the first term is simply $\kappa(n-1)^2$, as there are $\kappa(n-1)$ spanning trees in each F_{n-1}. Let us label the second term $\kappa'(n)$, we claim

$$\kappa'(n) = 2\kappa(n-1)\kappa'(n-1)$$

This claim is easy to prove. The first term comes from symmetry reasons. Let us assume that we initially span the left F_{n-1} of F_n, we have $\kappa(n-1)$ ways to do that. Once this F_{n-1} is spanned, and since the bounding edge needs to be used, it is obvious that the initial and end node of the right F_{n-1} will be connected via a path. That means that the number of spanning trees left to count is precisely $\kappa'(n-1)$, as this counts the number of spanning trees in F_{n-1} where the initial and end nodes are connected, hence the formula above. Therefore,

$$\kappa'(n) = 2^n \prod_{i=1}^{n} \kappa(n-i),$$

such that

$$\kappa(n) = \kappa(n-1)^2 + 2^n \prod_{i=1}^{n} \kappa(n-i),$$

or alternatively

$$\kappa(n-1) - \kappa(n-2)^2 = \kappa'(n-1) = 2^{n-1} \prod_{i=2}^{n} \kappa(n-i).$$

Then the following holds

$$\begin{aligned}
\kappa(n) &= \kappa(n-1)^2 + 2\kappa(n-1)2^{n-1} \prod_{i=2}^{n} \kappa(n-i) \\
&= \kappa(n-1)^2 + 2\kappa(n-1)[\kappa(n-1) - \kappa(n-2)^2] \\
&= 3\kappa(n-1)^2 - 2\kappa(n-1)\kappa(n-2)^2
\end{aligned}$$

n	$\kappa(n)$
1	3
2	21
3	945
4	1845585

Table 5: Number of spanning trees in F_n.

□

5.3.4 Largest eigenvalues $\lambda_{\max}$ for F_n^k

In this section we start by focusing on the set of graphs generated via the concatenation rule $\oplus$ as introduced in Definition 5.2.3. We initially are interested in exploring the role of k in the largest eigenvalue of the adjacency matrix. To begin with, we fix n and explore (numerically) how $\lambda_{\max}(F_n^k)$ changes as we increase k. We have calculated $\lambda_{\max}(F_n^k)$ for $1 \leq n \leq 4$ and $1 \leq k \leq 10$. Results are shown in Figure 5.11. After a transient growth, we notice that for each n the $\lambda_{\max}$ appears to converge to a finite value as k increases. This observation can be made rigorous:

Theorem 5.3.6. *Let $n > 0$ be fixed and consider the graph F_n^k as k increases. Then the largest eigenvalue of its adjacency matrix converges as $k \to \infty$.*

Proof. Recall that the largest eigenvalue is bounded by the largest degree of the graph hence in our case, $\lambda_{\max}(F_n^k) \leq d_{\max}$. Now, the node with the largest degree in F_n^2 is the central node, which by construction inherits the edges from the left and right boundary nodes in F_n. These boundary nodes have degree $n+1$, hence

$$d_{\max}(F_n^2 \equiv F_n \oplus F_n) = 2(n+1).$$

Adding additional copies of F_n does not change the maximum degree, because only one of the boundary nodes in F_n^k will have their degree increased from $n+1$ to $2(n+1)$. In other words, the node with largest degree is maintained constant as new motifs are concatenated. Therefore $\lambda_{\max}(F_n)$ is bounded from above. Furthermore, as a conse-

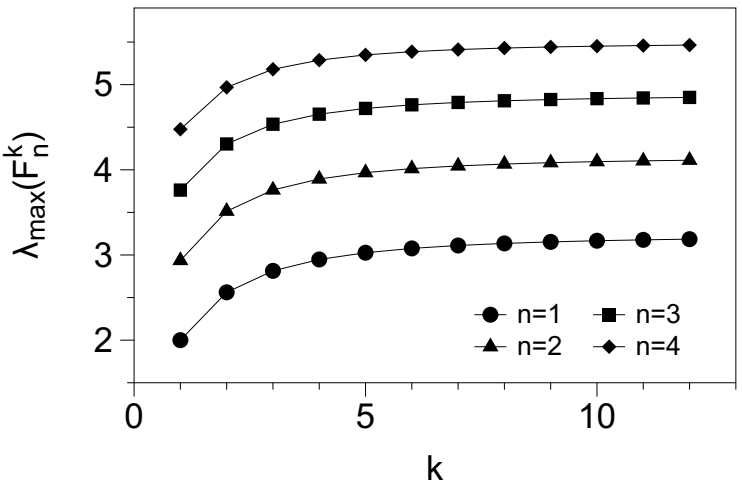


Figure 5.11: Plot of $\lambda_{\max}(F_n^k)$ for fixed values of n as a function of k. In each case the largest eigenvalue converges to a value independent of k, a result proved in Theorem 5.3.6.

quence of Cauchy's Interlacing Theorem we have that $\lambda_{\max}(F_n^{k+1}) \geq \lambda_{\max}(F_n^k)$. Therefore $\lambda_{\max}(F_n^k)$ is an increasing sequence in k, bounded above, hence converges. □

We have now understood the dependence of $\lambda_{\max}(F_n^k)$ on k, and we are now in a position to discuss a general expression for the largest eigenvalue in the general case of F_n^k. This is a two-parameter discrete function

$$\lambda_{\max}(F_n^k) : n, k \in \mathbb{N}^+ \to \mathbb{R}^+$$

We summarise the bounds for $\lambda_{\max}$ for general n and k in the following proposition:

Proposition 5.3.7. Consider the graph F_n^k, where $n \geq 0, k \geq 2$ (the case $k = 1$ reduces to F_n). Then the following hold:

(a) $d_{\max}(F_n^k) = 2(n+1)$ (independent of k).

(b) $\bar{d}(F_n^k) = \dfrac{2E_n^k}{V_n^k} = \dfrac{2k(2^{n+1}-1)}{k2^n+1}$ (asymptotically independent of k).

(c) $\max\{\sqrt{d_i d_j} : 1 \leq i, i \leq n, v_i v_j \in E\} = 2n+2$ (independent of k).

Proof. Proposition (a) comes from Theorem 5.3.6

Proposition (b) comes from Proposition 5.2.4 along with the fact that the average degree of a graph is twice the number of edges divided by the number of nodes

Proposition (c) is trivially proved by observing that for $k \geq 2$, F_n^k has multiple nodes with maximum degree; these nodes are always connected and have degree $2(n+1)$. □

These results provide the following bounds, which are equivalent to the bounds to Equations (5.3) to (5.8) but in the general case where $k \geq 2$:

$$\lambda_{\max}(F_n^k) \leq 2(n+1), \tag{5.12}$$

$$\bar{d}(F_n^k) = \frac{2k(2^{n+1}-1)}{k2^n+1} \leq \lambda_{\max}, \text{if } n < 7, \tag{5.13}$$

$$\sqrt{2(n+1)} \leq \lambda_{\max}, \text{if } n \geq 7. \tag{5.14}$$

Additionally, we were able to estimate the number of walks (as explained in Section 5.3.5) in the case of F_n^k, yielding:

$$a(n,k) = (320 \cdot 2^n - 4n^3 - 48n^2 - 196n - 312) + (k-2)(160 \cdot 2^n - 16n^2 - 88n - 152)$$

$$b(n,k) = 24 \cdot 2^n - 2n^2 - 12n - 22 + (k-1)(24 \cdot 2^n - 8n - 20)$$

$$c(n,k) = 4k \cdot 2^n - 2k$$

$$d(n,k) = k \cdot 2^n + 1.$$

5.3.5 Walk functions

In Section 5.3.2.1 we use the result $\|(\mathbf{A}_n)^2\|_\infty = 2n^2 + 2$, and we prove it here. Note that $\|(\mathbf{A}_n)^2\|$ is the maximum of the number of 2-walks originating at a node, over all the nodes. It is clear that this node is the central node, which we will call v_c, which has degree $2n$ (as we prove in Proposition 5.3.3). Also note that to count the number of 2-walks originating from v_c, we can count the total degree of the neighbours of v_c. We can observe that apart from the boundary (left and right) nodes, which have degree $n+1$, the degrees of the neighbours of the v_c are $2, 4, 6, \ldots, 2(n-1)$ (and these are

counted twice). Summing up all these degrees we have

$$\|(\mathbf{A}_n)^2\| = 2(n+1) + \sum_{k=1}^{n-1} 2k = 2n^2 + 2.$$

5.3.5.1 Coefficients

In Section 5.3.2.3 we defined $a(n)$, $b(n)$, $c(n)$ and $d(n)$ to be the total number of 3-walks, 2-walks, 1-walks and 0-walks respectively. We state that

$$b(n) = 24 \cdot 2^n - 2n^2 - 12n - 22$$

$$c(n) = 4 \cdot 2^n - 2$$

$$d(n) = 2^n + 1.$$

Observe that the number of 0-walks, $d(n)$, is the number of nodes, which is equal to $2^n + 1$. The number of 1-walks, $c(n)$, is twice the number of edges, and is equal to $2(2^{n+1} - 1) = 4 \cdot 2^n - 2$.

Reaching the formula for $b(n)$ is a little trickier. First define I_n^1 to be the $n \times n$ matrix with zeros everywhere, except a 1 in the top right entry. Similarly define ${}_1I_n$ to to be the $n \times n$ matrix with zeros everywhere, except a 1 in the bottom left entry. Notice that

$$\mathbf{A}_n = \mathbf{A}_{n-1}^2 + I_n^1 + {}_1I_n, \tag{5.15}$$

where $\mathbf{A}_{n-1}^2$ is the adjacency matrix of $F_{n-1}^2 = F_{n-1} \oplus F_{n-1}$ as in Definition 5.2.3. The quantity we wish to find is $\sum_{i,j}(\mathbf{A}_{n-1}^2)^2$. There is an unfortunate use of notation, so one must remember that the superscript next to the letter $\mathbf{A}$ corresponds to the concatenation rule, and if we wish to square the matrix, we will explicitly use brackets. We plot a visualisation of the matrix $(\mathbf{A}_{n-1}^2)^2$ in Figure 5.12. The source of the top left and bottom right blocks $(\mathbf{A}_{n-1})^2$ should be clear by studying the form of the matrix $\mathbf{A}_{n-1}^2$. A matrix $\mathbf{C}$ appears in the top right (with its transpose in the bottom left), and has size $2^{n-1} \times 2^{n-1}$. The origin of this matrix is not immediately clear, however it is the matrix obtained when the central row vector of $\mathbf{A}_{n-1}^2$ hits itself under squaring. The

vector has sum $2n$ (recall the degree of the central vertex), but only half of the vector hits itself when creating $\mathbf{C}$, so we will truncate this vector and only consider the first 2^{n-1} values of this central vector, and we will call it v_c. Thus $\mathbf{C}$ is the matrix where the nth row vector is v_c if the nth value of v_c is 1, and is zero otherwise (or rather, a vector of zeros of length 2^{n-1}), or equivalently

$$\mathbf{C} = \underbrace{(v_c^\intercal | v_c^\intercal | \cdots | v_c^\intercal)}_{2^{n-1}}.$$

This is significantly more easy to follow if one squares, with a pen and paper, the matrix $\mathbf{A}_{n-1}^2$. The sum of v_c is n, so we have that $\sum_{i,j} \mathbf{C} = \sum_{i,j} \mathbf{C}^\intercal = n^2$. Going back to Equation (5.15) we have

$$\begin{aligned}(\mathbf{A}_n)^2 &= (\mathbf{A}_{n-1}^2 + I_n^1 + {}_1I_n)^2 \\ &= (\mathbf{A}_{n-1}^2)^2 + \mathbf{A}_{n-1}^2 \cdot I_n^1 + \mathbf{A}_{n-1}^2 \cdot {}_1I_n + I_n^1 \cdot \mathbf{A}_{n-1}^2 \\ &+ (I_n^1)^2 + I_n^1 \cdot {}_1I_n + {}_1I_n \cdot \mathbf{A}_{n-1}^2 + {}_1I_n \cdot I_n^1 + ({}_1I_n)^2.\end{aligned}$$

We then sum this quantity over i and j, the first term gives a contribution of twice that of $\mathbf{A}_{n-1}^2$ (see Figure 5.12) and twice the sum of $\mathbf{C}$. The terms involving $\mathbf{A}_{n-1}^2$ and either I_n^1 or ${}_1I_n$ give us a contribution of n, as these vectors extract the top, bottom, left and right row/column vectors of $\mathbf{A}_{n-1}^2$ and these have sum n (recall the degree of the left or right boundary nodes is $(n-1)+1 = n$). The terms $(I_n^1)^2$ and $({}_1I_n)^2$ have sum 0 but the terms $I_n^1 \cdot {}_1I_n$ and ${}_1I_n \cdot I_n^1$ have sum 1 each. Putting this together we have

$$\sum_{i,j}(\mathbf{A}_n)^2 = 2 \cdot \sum_{i,j}(\mathbf{A}_{n-1})^2 + 2n^2 + 4n + 2,$$

and writing $\sum_{i,j}(\mathbf{A}_n)^2 = b(n)$ we have a recurrence relation

$$b(n) = 2 \cdot b(n-1) + 2n^2 + 4n + 2.$$

We have that $b(0) = 2$, hence this can be solved and we get

$$b(n) = 24 \cdot 2^n - 2n^2 - 12n - 22,$$

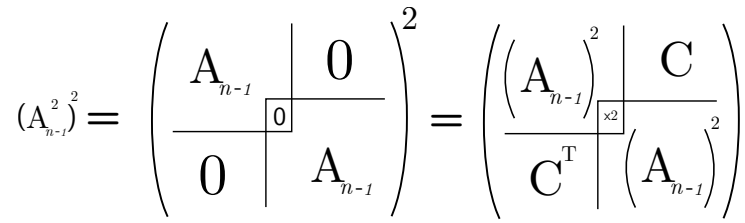


Figure 5.12: Diagram of the matrix representation of $(F_{n-1}^2)^2$ in terms of the matrices $\mathbf{A}_{n-1}$. The middle entry is the sum of the bottom right entry and top left entry of $(\mathbf{A}_{n-1})^2$ (or equivalently twice the either entry as the matrix is symmetric). An extra matrix $\mathbf{C}$ appears in the top right and bottom left blocks, whose entries sum to n^2, as explained in the text.

completing the proof. Constructing a similar proof for the 3-walks $a(n)$ could be possible but we were not able to. However, we can guess that $a(n)$ is of the form $m \cdot 2^n + g(n)$ where $g(n)$ is a polynomial in n and m is an integer. Calculating $a(n)$ directly for enough values of n we can estimate the coefficients (noting that it's an exact fit for as many values of n that we could calculate), and again we find integer coefficients with

$$a(n) = 160 \cdot 2^n - 4n^3 - 30n^2 - 104n - 158.$$

We can extend this analysis by trying to create formulas for the 0, 1, 2, and 3-walks of F_n^k which we define as $d(n,k)$, $c(n,k)$, $b(n,k)$ and $a(n,k)$ respectively. We immediately get $d(n,k)$ and $c(n,k)$ from the definitions. We can estimate the other two by proceeding with the same method as for $a(n)$ by guessing the general form of $b(n,k)$ and $a(n,k)$ to be $m \cdot 2^n + g(n) + k \cdot h(n)$ where m is an integer and $g(n)$ and $h(n)$ are polynomials. The new contribution of $h(n)$ comes from adding k copies of a certain number of walks. This yields a conjecture:

$$a(n,k) = (320 \cdot 2^n - 4n^3 - 48n^2 - 196n - 312) + (k-2)(160 \cdot 2^n - 16n^2 - 88n - 152)$$

$$b(n,k) = 24 \cdot 2^n - 2n^2 - 12n - 22 + (k-1)(24 \cdot 2^n - 8n - 20)$$

$$c(n,k) = 4k \cdot 2^n - 2k$$

$$d(n,k) = k \cdot 2^n + 1.$$

5.3.6 The complete spectrum of F_n^k: tridiagonal n-block Toeplitz matrices.

We now turn our attention to the adjacency matrices of F_n^k and their particular form. As a preamble, observe that the concatenation operation $\oplus$ that generates F_n^k from F_n is in some sense 'close' to a direct sum. We recall that the direct sum of a matrix $\mathbf{A}$ with itself is the matrix formed by placing $\mathbf{A}$ as two non-overlapping diagonal blocks. The eigenvalues of the direct sum of two copies of the same matrix $\mathbf{A}$ are just the eigenvalues of $\mathbf{A}$ (with twice the multiplicity in each case). If we 'approximate' $\oplus$ as just being the direct sum operation, then trivially the eigenvalues of F_n^k would be the same as the eigenvalues of F_n (with multiplicity). In particular, $\lambda_{\max}$ would be fully independent of k. Of course, $\oplus$ is not a direct sum, however $\lambda_{\max}(F_n^k)$ is independent of k in the limit $k \to \infty$. With a bit of hand-waving, we could say that the larger n, the 'closer' $\oplus$ is to a direct sum and therefore the more independent the spectrum is from k.

We start now our analysis by fixing n and letting k increase. For $n = 0$, F_0^k is trivially a 2-regular chain whose adjacency matrix $\mathbf{A}_0^k$ whose structure is tridiagonal Toeplitz:

$$\mathbf{A}_0^k = \begin{bmatrix} 0 & 1 & 0 & & 0 & 0 \\ 1 & 0 & 1 & \ddots & & 0 \\ 0 & 1 & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & 1 & 0 \\ 0 & & \ddots & 1 & 0 & 1 \\ 0 & 0 & & 0 & 1 & 0 \end{bmatrix}$$

Accordingly, through direct calculation of the adjacency matrix, we can express the spectrum in closed form

$$\mathrm{spec}(F_0^k) = \{2\cos\frac{h\pi}{k+2},\ h = 1, \ldots, k+1\}.$$

A plot of this spectrum for $k = 2^{10}$ is shown in Figure 5.13. In particular, as $\cos(x)$

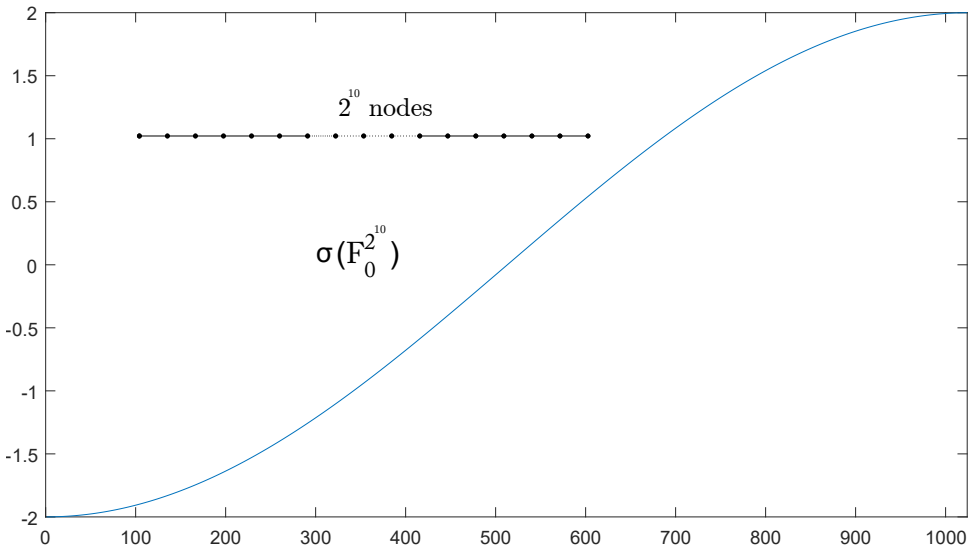


Figure 5.13: The spectrum of $F_0^{2^{10}}$.

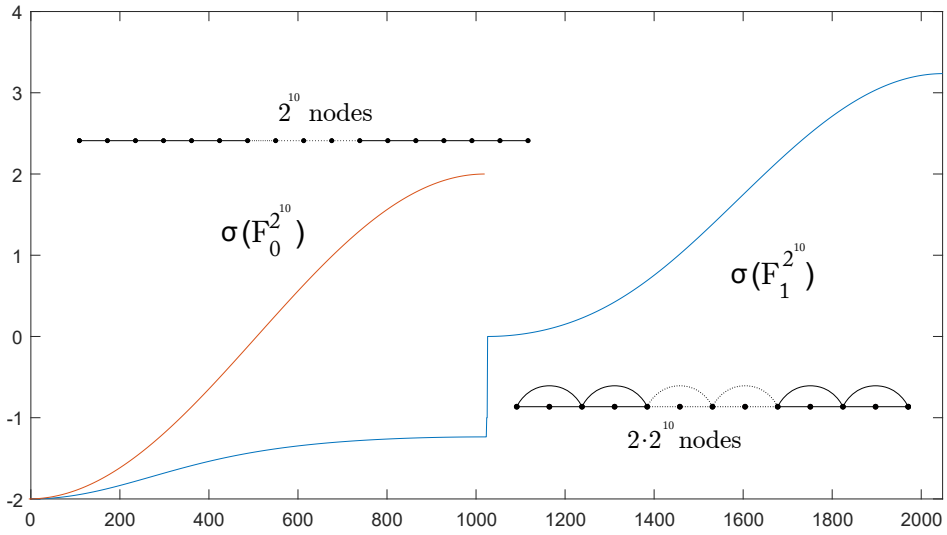


Figure 5.14: The spectrum of $F_0^{2^{10}}$ and $F_1^{2^{10}}$.

monotonically decreases on $[0, \pi]$, the largest value is found for $h = 1$ and thus

$$\lambda_{\max}(F_0^k) = 2\cos\frac{\pi}{k+2},$$

hence we have that $\lim_{k\to\infty} \lambda_{\max}(F_0^k) = 2$.

For $n = 1$, F_1^k, the adjacency matrix is no longer tridiagonal Toeplitz, however it can be

expressed as a tridiagonal *block* Toeplitz matrix of the shape

$$\begin{pmatrix} a & b & 0 & \dots & \\ b^T & a & b & 0 & \dots \\ 0 & b^T & a & b & 0 \end{pmatrix}$$

where

$$a = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ b = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},\ b^T = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

This is a special type of tridiagonal block Toeplitz matrix. In general, if we look at the adjacency matrix $\mathbf{A}_n^k$ associated to F_n^k, there exists a self-similar process underlying the construction of $\mathbf{A}_n^k$ in terms of $\mathbf{A}_{n-1}^k$. For instance, $\mathbf{A}_0^k$ is just a tridiagonal Toeplitz matrix with null diagonal elements. Now, $\mathbf{A}_1^k$ is not tridiagonal nor Toeplitz as we have seen, but we recover a tridiagonal Toeplitz shape if we consider blocks 2×2 as the elements of this new matrix, or equivalently $\mathbf{A}_1^k$ is a tridiagonal block Toeplitz matrix. Similarly, $\mathbf{A}_2^k$ is no longer a tridiagonal block Toeplitz matrix, but if we consider that the elements of $\mathbf{A}_2^k$ are blocks of blocks (2×2 matrices whose elements are in turn blocks), then in the structure of $\mathbf{A}_2^k$ is again tridiagonal Toeplitz (we may call it tridiagonal superblock, or 2-block Toeplitz). For instance, the structure of $\mathbf{A}_2^k$ can be expressed as

$$\begin{pmatrix} A & B & 0 & \dots & \\ B^T & A & B & 0 & \dots \\ 0 & B^T & A & B & 0 \end{pmatrix}$$

where

$$A = \begin{pmatrix} a & b \\ b^T & a \end{pmatrix},\ B = \begin{pmatrix} c & 0 \\ b & 0 \end{pmatrix},\ B^T = \begin{pmatrix} c & b^T \\ 0 & 0 \end{pmatrix},\ c = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

This process can be applied iteratively and hence we can show that $\mathbf{A}_n^k$ has a tridiagonal n-block Toeplitz structure. In this case, an n-block is equivalent to a $2^n \times 2^n$ block. In other words, a tridiagonal n-block Toeplitz matrix is equivalent to a tridiagonal block

Toeplitz matrix where each block is a $2^n \times 2^n$ matrix. We weren't able to find such shape in the literature but we speculate that the set of symmetries present in the recursive building of $\mathbf{A}_n^k$ could be exploited to extract properties about its spectrum. Additionally, in Figure 5.14 we plot the spectrum of $F_0^{2^{10}}$ and $F_1^{2^{10}}$. For the spectrum of $F_1^{2^{10}}$, we notice two distinct curves separated by a discontinuity, with a length of approximately 2^{10} each. Each of these two curves look like appropriate re-scalings of the spectrum of $F_0^{2^{10}}$. The same pattern can be observed in $F_n^{2^{10}}$ for $n = 2, 3, 4$ with the spectrum of $F_n^{2^{10}}$ having 2^n distinct curves, each separated by a jump. We conjecture that for a fixed k, the spectrum of F_n^k consists of 2^n distinct curves.

Finally, in Figure 5.15 we plot the complete point spectrum of F_{10}, and compare it with F_n^k with the same number of nodes: F_8^4, F_6^{16}, F_4^{64} and F_2^{256}. We can see how for small n the distinct curves are very obviously separated by discontinuities, and these smear out as n increases. The spectrum seems to converge to a somewhat universal shape. We conjecture that this self-similar process is reminiscent of the recursive way of building the n-block tridiagonal Toeplitz adjacency matrices, and we leave this as an open problem.

5.3.7 Determinant of Feigenbaum Graphs

We close this section on the properties of F_n^k by exploring the determinant of F_n^k, which is defined as the determinant of the adjacency matrix $\mathbf{A}_n^k$. We outline and prove the following theorem:

Theorem 5.3.8. *The determinants of F_n^k satisfy*

$$det(F_n) = \begin{cases} -1, & n = 0 \\ 2, & n = 1 \\ -2, & \forall n \geq 2. \end{cases}$$

Moreover, for $k \geq 1$ and $n \geq 2$ we have

$$det(F_n^k) = -2k$$

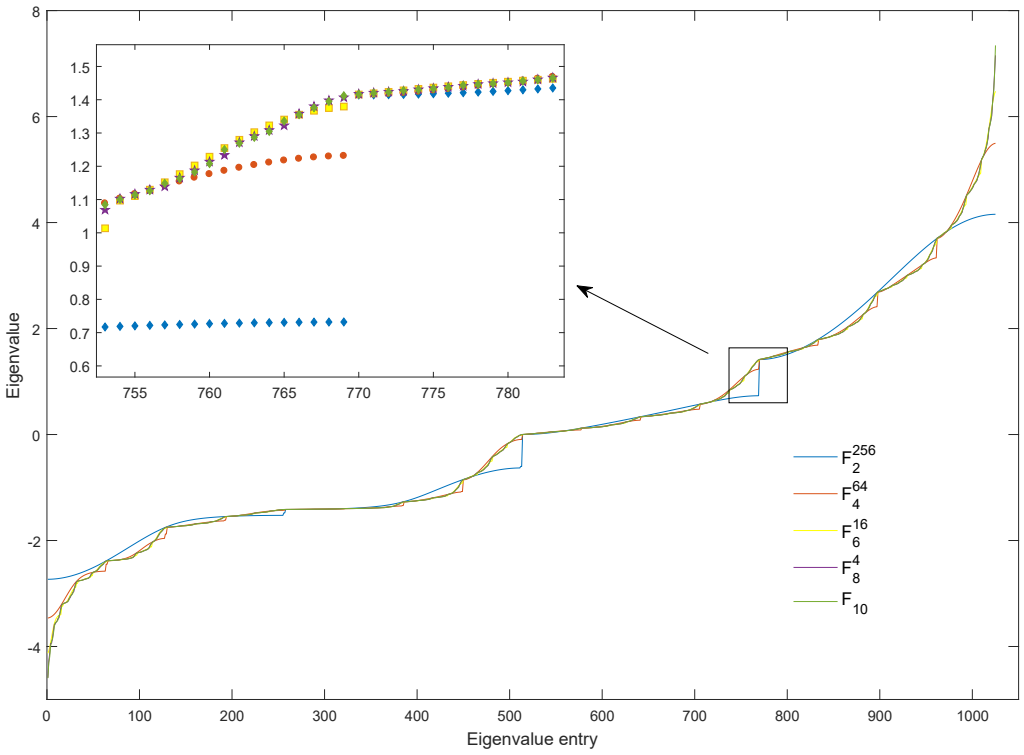


Figure 5.15: Rescaled curves for F_n^k, where we show that, when comparing graphs with the same number of nodes, the spectrum collapses to a universal shape.

Proof. For $k = 1$, we can directly calculate $det(F_0)$ and $det(F_1)$. To push beyond this is a little bit more difficult. It is too tricky to directly calculate the determinant of the adjacency matrices, despite them having a recursive form. We follow a graph theoretical proof, for which we will have to (i) define the concept of spanning elementary subgraph in Definition 5.3.9, (ii) state (without proof) a theorem of Harary [92] (Theorem 5.3.10), and then (iii) state and prove two lemmas (Lemmas 5.3.11 and 5.3.12. The rest of the proof will then follow easily. We start by defining a *spanning elementary subgraph* [91]:

Definition 5.3.9 (Spanning Elementary Subgraph)**.** An elementary subgraph is a simple subgraph, each component of which is regular and has degree 1 or 2, i.e., each component is either a single edge or a cycle. A spanning elementary subgraph (SES) of a graph G is an elementary subgraph which contains all vertices of G.

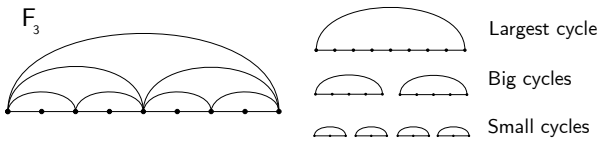


Figure 5.16: Diagram showing F_3 with the relevant cycles. The largest cycle is taken as the Hamiltonian path along with the edge connecting the first and last vertices. The small cycles are always the subgraph K_3, and the rest of the cycles are labelled as big cycles.The big and small cycles are taken by following the Hamiltonian path but taking an edge back to the starting node.

We now make use of the following Theorem:

Theorem 5.3.10 (Harary 1962; [92])**.** *Let A be the adjacency matrix of a graph G, let v be the number of vertices, e the number of edges and l the number of components. Then*

$$det(A) = \sum_{H}(-1)^{r(H)}2^{c(H)} \tag{5.16}$$

where the summation is over all spanning elementary subgraphs H of G, $r(H) = v - l$ is the rank of H and $c(H) = e - v + l$ is the co-rank.

We note that the co-rank of an elementary subgraph is just the number of cycles in the graph. Our task is thus to find all the spanning elementary subgraphs with their corresponding ranks and co-ranks.

Lemma 5.3.11. *We have only two configurations for elementary subgraphs (for $n \geq 2$). The first is just the cycle containing all vertices (shown as the "largest cycle" in Figure 5.16) and the second is any other cycle from Figure 5.16 with the remaining nodes connected by single edges*

Proof. Without loss of generality we consider the structure of F_3 only. Because of the recursive property of the Feigenbaum graphs, all the arguments used here can be applied directly to any F_n with $n \geq 2$.

The largest cycle (created by taking only the outside edges of the outerplanar graph) is

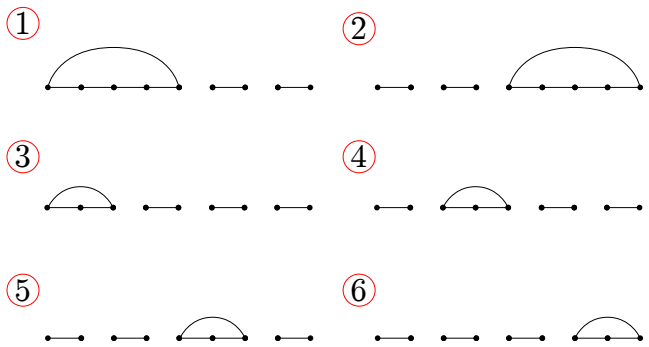


Figure 5.17: Six other configurations for elementary spanning subgraphs of F_3 (other than the largest cycle).

shown in Figure 5.16. The co-rank of this subgraph is 1 (since the only component is a cycle, and the co-rank is the number of cycles) and the rank is even as the number of vertices in any Feigenbaum graph is always odd, and we only have one component). Thus this subgraph contributes 2 in the sum (Equation (5.16)) of the determinant of F_3. This is true for any n, i.e., the largest cycle contributes 2 towards the determinant.

We can construct other spanning elementary subgraphs by taking any other cycle (big or small as in Figure 5.16) and joining the remaining bottom edges. As such cycles always contain an odd amount of vertices, we are always left with an even amount of vertices on the bottom, which permits us to join the rest of the vertices with single edges. Such spanning elementary subgraphs, for F_3 are shown in Figure 5.17.

We stipulate the following: *we cannot have more than one cycle in any elementary subgraph*. This is because if we take two cycles in our elementary subgraph, we will be left with an odd amount of vertices. An odd amount of vertices cannot be joined only by single edges, thus we would require another cycle to give us an even amount of vertices. However, because of the construction of the Feigenbaum graphs, this will leave us with an odd amount on either side of one of the cycles, and this process repeats until we are left with a single node that cannot by introduced in to any spanning elementary subgraph.

By the same reasoning, any other cycles considered which are not listed in Figure 5.16

(for example in F_3, taking the triangle formed by the 1st, 3rd and 5th nodes) will again leave us with an odd amount of nodes, the first of which (by ordering the remaining nodes and numbering them left to right starting with 1) can only be connected by a single edge to the next node. This process repeats until we are left with a single node that cannot be introduced in to any spanning elementary subgraph.

Thus each spanning elementary subgraph contains only one of our big or small cycles, and each big or small loop corresponds to exactly one spanning elementary subgraph. The co-rank of all elementary subgraphs of all F_n is therefore equal to 1. This concludes the proof of Lemma 5.3.11. □

Lemma 5.3.12. *Spanning elementary subgraphs consisting of a big cycle have $r(H)$ equal to an even number. SES's consisting of a small cycle have $r(H)$ equal to an odd number.*

Proof. The number of vertices of F_n is $2^n + 1$. The number of vertices in one of the big or small cycles is $2^m + 1$ where $1 \leq m < n$. The number of vertices remaining when we add a cycle component (a loop) to a SES is $2^n + 1 - (2^m + 1) = 2^n - 2^m$ and the number of single edge components is

$$\frac{2^n - 2^m}{2} = 2^{n-1} - 2^{m-1}$$

which is even if and only if $m \neq 1$, i.e., only for big cycles. After adding in the single cycle component, we have that the number of components is odd only when $m \neq 1$, i.e., only for big cycles. Recall that $r(H) = v - l$, where l is the number of components. The number of vertices is odd, hence we now have $r(H)$ is even if and only if $m \neq 1$, i.e., only for big cycles. This concludes the proof of Lemma 5.3.12. □

Going back to our formula in Equation (5.16), we have $c(H) = 1$, therefore

$$det(A) = \sum_H 2(-1)^{r(H)}$$

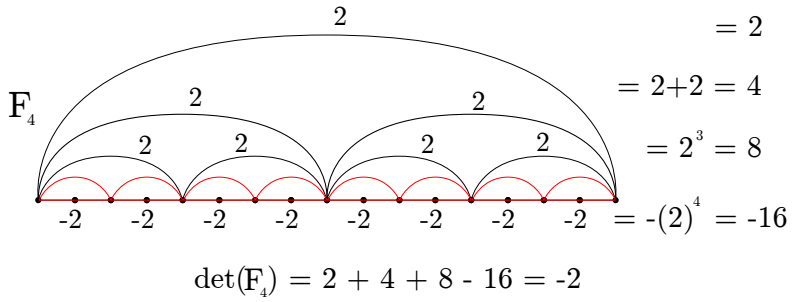


Figure 5.18: Diagram showing each cycle's contribution to $det(F_n)$.

In Lemma 5.3.12 we proved that $r(H)$ is even for SES containing big cycles and odd for small. Using simple combinatoric arguments shown in Figure 5.18, we have for $n \geq 2$,

$$det(F_n) = -2^n + \sum_{j=1}^{n-1} 2^j = -2$$

For $F_n^{k>1}$ the big and small cycles give slightly different contributions to the determinant, similar to Figure 5.18, however it is easy to see that each additional concatenation adds another contribution of -2 to the determinant, hence k copies of F_n give us a determinant of $-2k$, giving us:

$$det(F_n^k) = -2k,$$

which concludes the proof of Theorem 5.3.8. □

Remark. It can also be checked that, using similar arguments to the proof of Theorem 5.3.8, $det(F_1^k) = (-1)^{k+1}2k$ and that for $n = 0$:

$$det(F_0^k) = \begin{cases} 0 & k \text{ even} \\ (-1)^{\frac{k+1}{2}} & k \text{ odd.} \end{cases}$$

5.4 $\mu > \mu_\infty$: Spectral properties of chaotic Feigenbaum graph ensembles

In this section we explore Feigenbaum graphs in the region $\mu > \mu_\infty$. As discussed in Section 5.2.3, for a given $\mu < \mu_\infty$, and for a given series size N, the resulting Feigenbaum

graph was unique because the order in which the trajectory visits the stable branches of the periodic attractor is unique (indeed, it is universal for all unimodal maps, not just the logistic map, so F_n^k are universal [84]). However, for $\mu > \mu_\infty$ this is no longer the case: for a specific μ, each initial condition will generate a priori a different chaotic trajectory, and hence a different Feigenbaum graph. Since in this case n and k do not apply, we use the notation $F(\mu, N)$ to describe the ensemble of Feigenbaum graphs associated to a trajectory of size N (so the corresponding HVG has N vertices) generated by the logistic map with parameter μ.

5.4.1 Self-averaging properties of $\lambda_{\max}$

We start by exploring the self-averaging properties of the ensembles of Feigenbaum graphs. First, we fix $\mu = 4$ (fully developed chaos) and extract an ensemble of 100 time series of series with $N = 2^q$ for $q = 10, \ldots, 15$, each generated with a different initial condition. For each series, we then extract its Feigenbaum graph and calculate $\lambda_{\max}$. For each time series size N, we compute the mean and standard deviation of the ensemble of $\lambda_{\max}$. To assess whether this quantity self-averages as N increases [93], in the left panel of Figure 5.19 we plot the relative variance R_λ as a function of N, defined as

$$R_\lambda(N) = \frac{\langle \lambda_{\max}^2 \rangle - \langle \lambda_{\max} \rangle^2}{\langle \lambda_{\max} \rangle^2},$$

where the average $\langle \cdot \rangle$ is performed over the ensemble of realisations.

We observe that this quantity decreases with N, certifying that, for $\mu = 4$, the largest eigenvalue is a self-averaging quantity. This means that with regards the largest eigenvalue, a *typical* realisation of $F(\mu = 4, N)$ provides a faithful representation of the ensemble. Moreover, the relative variance scales as a power law $R_\lambda(N) = cN^z$ with $z \approx -0.221$, hence the system is *weakly* self-averaging (because we have $-1 < z < 0$).

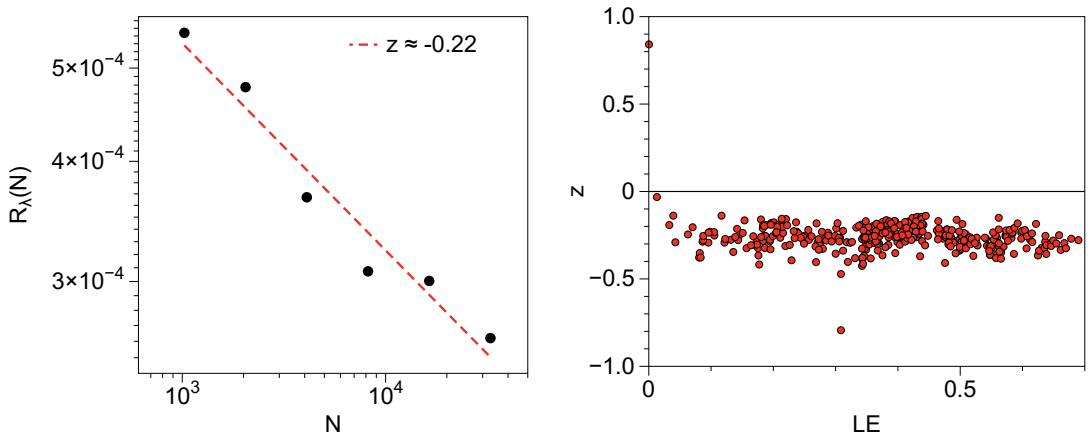
A similar analysis[1] is performed now for the whole range of values of $\mu > \mu_\infty$ for which the Lyapunov exponent (LE) is positive (i.e., we discard periodic windows). In each case,

[1] Power law functions are fitted here using a least squares algorithm.

a power law fit $R_\lambda(N) = cN^z$ is computed. In the right panel of Figure 5.19 we plot the estimated exponent $z(\mu)$. In most of the cases we find that the system remains weakly self-averaging. There is only one exception for this otherwise general behaviour: for a specific value of μ only slightly above μ_∞ (LE $\approx 0+$) we find that $z > 0$, i.e., the relative variance increases with N. This anomalous behaviour can be explained as follows: in the onset of chaos $\mu = \mu_\infty$, the Feigenbaum graph ensemble is still degenerate (i.e., only one unique configuration). As we enter into the chaotic region but remain very close to μ_∞, a trajectory of the map will visit what is known as a ghost of the attractor found in the accumulation point. In fact, the structure of a realisation of a Feigenbaum graph just above the accumulation point is very similar to the one found at the accumulation point with just a few additional 'chaotic' edges [84]. The existence of these edges is what allows the ensemble in this case to no longer be degenerate. Now, the number of these chaotic edges will proportionally increase when the series size N increases, simply because as N increases the trajectory will show additional deviations from the ghost attractor. Accordingly, the total number of possible configurations of the ensemble of Feigenbaum graphs very close to the accumulation point increases from essentially one (degenerate case) when N is small to many as N increases. As a by-product, the relative variance will necessarily increase as a function of N in this case, hence $z > 0$.

5.4.2 Searching spectral correlates of chaoticity

One of the main motivations that leads us to explore the largest eigenvalue of HVGs is that some research claims that this is an informative quantity for the 'complexity' of the associated time series (see for instance [30, 80–83]). If this was the case, we wonder if such quantity is able to quantify the 'degree of chaoticity' of a given (chaotic) time series. Within the realm of nonlinear time series analysis, a relevant property that quantifies how chaotic a system is the *sensitivity to initial conditions*, better described by the largest Lyapunov exponent of the system which accounts for the (exponential) separation rate of two initially nearby trajectories. For univariate time series extracted from a map $x_{i+1} = f(x_i)$, there is only one Lyapunov exponent LE, which can be estimated from a



(a) (b)

Figure 5.19: (a) Log-log Plot showing the relative variance $R_\lambda(N)$ as a function of size N, computed over an ensemble of 500 realisations of Feigenbaum graphs $F(4, 2000)$. The curve is fitted by a power law $R_\lambda(N) = cN^z$, where the best fit provides $z \approx -0.22$, suggesting the system is weakly self-averaging ($-1 < z < 0$). (b) Fitted exponent z for the range of values of μ for which $\text{LE}(\mu) > 0$. In most of the cases we find $-1 < z < 0$, confirming that the system is weakly self-averaging. For $\text{LE}(\mu) \approx 0+$ (which holds for μ only slightly above μ_∞), we find $z > 0$: this can be explained in terms of the ghost structure present in the graphs (see the text).

single (long) time series as [94]

$$\text{LE} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \log |f'(x_i)|$$

Thus, for each $\mu \in [\mu_\infty, 4]$ (sampled in steps of $\Delta\mu = 0.001$) we have generated a single trajectory, and computed both the LE and $\lambda_{\max}$. In Figure 5.20 we show the scatter plot of $\lambda_{\max}$ vs LE. Surprisingly, no obvious correlation emerges in this picture, which suggests that $\lambda_{\max}$ does not correlate to the sensitivity to initial conditions.

Does this mean that HVGs are not inheriting chaoticity properties, or that these are simply not inherited in $\lambda_{\max}$? As a matter of fact, previous works have shown that the HVGs *do* capture chaoticity, as $\text{LE}(\mu)$ is very well approached (from above) by suitable block-entropies of the Feigenbaum graph's degree sequence [95]. So the question is whether the spectral properties of these graphs are able to capture such properties. We do not have a definite answer for this, but let us comment that we have checked scatter plots similar to Figure 5.20 for other spectral properties, such as the graph's

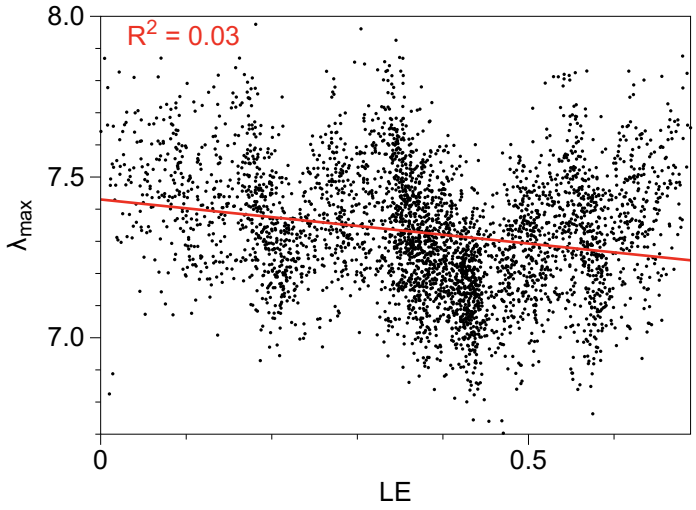
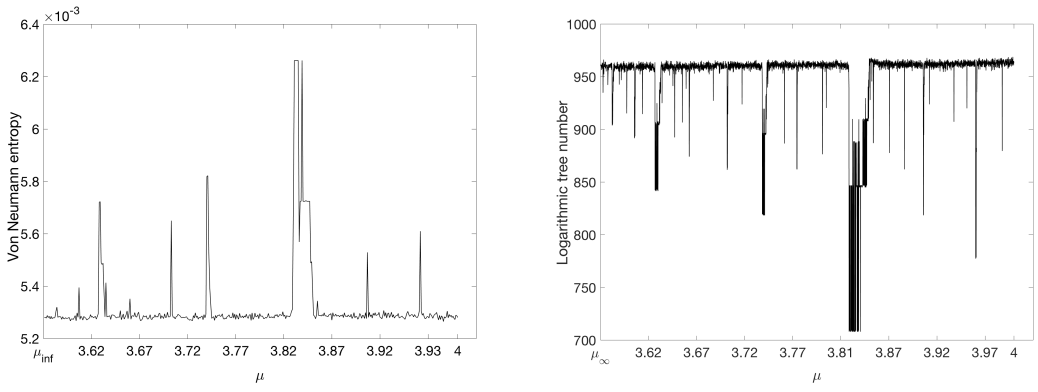


Figure 5.20: Scatter plot of the maximum eigenvalue $\lambda_{\max}$ of $F(\mu, 2000)$ vs the Lyapunov exponent $\text{LE}(\mu)$, for values of $\mu \in [\mu_\infty, 4]$ in steps of $\Delta\mu = 0.001$ (only positive values are selected to avoid periodic windows). No correlation emerges.

Von Neumann entropy [96] or the (logarithmic) tree number, with similarly unsuccessful results (see Figure 5.21 for details). Hence our partial conclusion is that spectral properties do not quantify different levels of chaoticity. The natural question is therefore: do these characterise chaos *at all*? To address this question, in the next and final section of the chapter we will make a systematic comparison between the spectral properties of Feigenbaum graphs associated to chaotic series and those of generic HVGs associated to random uncorrelated series.

5.4.3 Comparison with iid

In Section 5.4.2 we came to the conclusion that spectral properties don't seem to characterise (in a quantitative way) the chaoticity of the series. Hence the question: do they carry *qualitative* information, or on the contrary, spectral properties do not distinguish between chaotic series and random ones? If this was to be the case, the spectral properties shouldn't differ much from what we would find for random, uncorrelated series (iid).



(a) (b)

Figure 5.21: Von Neumann entropy (left) and logarithmic tree number (right) of the Feigenbaum graph associated to a time series of $N = 2000$ time steps extracted from a chaotic logistic map, as a function of the map's parameter μ. The two metrics are rather noisy at the ranges of μ where the map is chaotic and there is no obvious dependence with the Lyapunov exponent of the map, suggesting that these spectral properties do not capture the degree of chaoticity of the associated time series.

5.4.3.1 λ_{max}

First let us note that in [83] the authors explored whether $\lambda_{\max}$ could distinguish chaotic and random series, with interesting numerical evidence suggesting that chaos can be distinguished from an iid process under this lens. As a cautionary note, observe however that their analysis was based on estimating $d_{\max}$, as they claim that $\lambda_{\max} \approx \sqrt{d_{\max}}$ when $N \to \infty$. This is however not true in general (for a generic graph), and in the context of HVGs it is actually unknown. Also, they assumed that this quantity converged as the series size N increases, and numerically checked this in a small interval of N. Note, however, that analytical results [36] suggest that $d_{\max}$ is unbounded for both iid and chaotic processes as their degree distribution has an exponential tail (that is to say, in order to find a certain value for $d_{\max}$ one just needs to increase (exponentially) the series size N).

Does $\lambda_{\max}$ converge as $N \to \infty$? Since both iid and chaotic series are aperiodic, from Equation (5.1), we get $\bar{d} = 4$ in both cases. Furthermore, from [4, 36] it is known that degree distribution for both infinite iid and a chaotic process such as the logistic

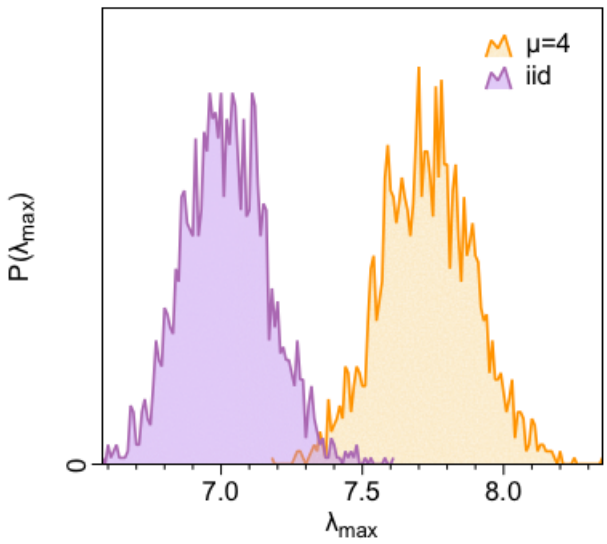


Figure 5.22: Ensemble histogram $P(\lambda_{\max})$ for $\mu = 4$ and i.i.d (the histogram has been smoothed).

map has an exponential tail, with $P(d) \sim \exp(-\gamma d)$. In particular, for $\mu = 4$ a good approximation is $\gamma_{\mu=4} \approx \log(4/3)$, whereas for an iid process the exponential distribution is exact and $\gamma_{\text{iid}} = \log(3/2)$, regardless of the marginal distribution (i.e., this holds for Gaussian iid, uniform iid, etc). Note, however that these expressions hold in the limit $N \to \infty$, where $d_{\max}$ is unbounded (although it grows rather slowly with N) in both cases, suggesting that $\lambda_{\max}$ is indeed unbounded in the limit $N \to \infty$. This is not unexpected, as $F(\mu, \infty)$ are not locally finite. For that reason, in order to assess whether $\lambda_{\max}$ can indeed distinguish chaos from iid, we shall analyse finite trajectories ($N < \infty$). $d_{\max}$ is therefore the largest possible degree of $F(\mu, N)$. Statistically speaking, we can state that $d_{\max}$ is only reached once in the whole graph, and therefore $d_{\max}$ should fulfil

$$P(d = d_{\max}) \cdot N = 1$$

A quick calculation yields

$$d_{\max} \sim \frac{\log N}{\gamma}$$

and according to Equations (5.4) and (5.5), we have for both iid and chaos:

$$\sqrt{d_{\max}} \leq \lambda_{\max} \leq d_{\max}$$

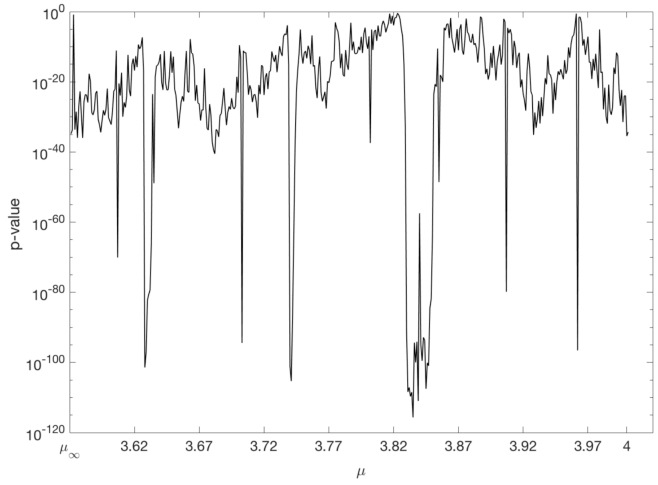


Figure 5.23: p-values for the 2-sampled t-test between $\langle \lambda^{\mu}_{max} \rangle$ and $\langle \lambda^{iid}_{max} \rangle$, estimated by averaging each case over 50 realisations of white uniform noise and the logistic map, for varying values of μ. The plot is in semi-log scale, revealing extremely low p-values for most of the range.

Interestingly, the difference between the chaotic case ($\mu > \mu_\infty$) and the random case (iid) is evident in $d_{\max}$:

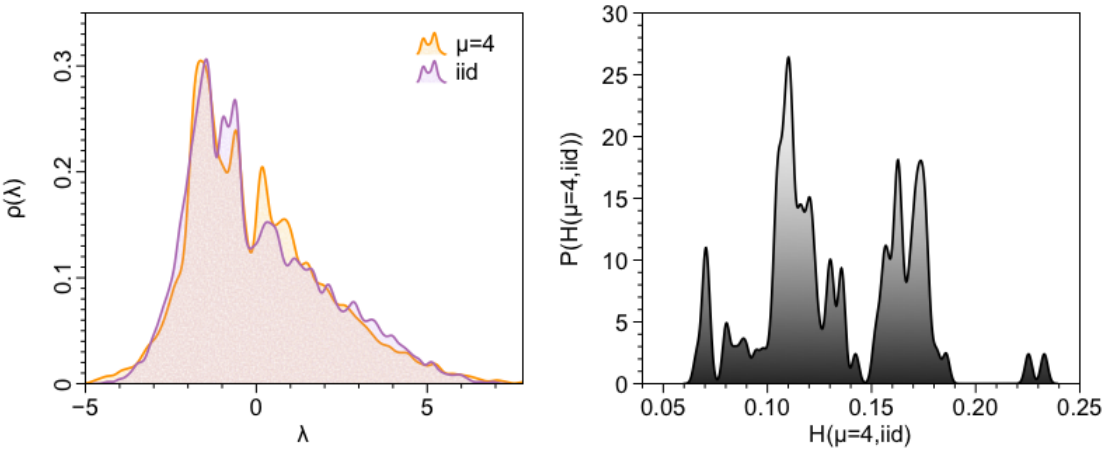
$$d^{\mu}_{\max} = \frac{\gamma_\mu}{\gamma_{\text{iid}}} d^{\text{iid}}_{\max}$$

which for $\mu = 4$ becomes

$$d^{\mu=4}_{\max} \approx 1.4 \cdot d_{\max}\text{iid}$$

We fix $N = 2000$ and compute $\lambda_{\max}$ for iid (where all random variables $x_t \sim \mathrm{U}(0,1)$) and $\mu = 4$ over 2000 realisations . We plot the resulting histograms in Figure 5.22, finding $\langle \lambda^{\text{iid}}_{\max} \rangle = 7.01 \pm 0.15$, and $\langle \lambda^{\mu=4}_{\max} \rangle = 7.73 \pm 0.16$. The two quantities are clearly different.

We now assess whether $\lambda_{\max}$ of an ensemble of logistic maps is systematically different than the same quantity obtained from iid. To do this, we consider all values of μ for which $\textsc{le}(\mu) > 0$ and for each of these values, we have performed a 2-sampled t-test between $\langle \lambda^{\mu}_{\max} \rangle$ and $\langle \lambda^{\text{iid}}_{\max} \rangle$, and obtained a p-value for each test. We systematically find very small p-values, concluding that $\langle \lambda_{\max} \rangle$ can distinguish time series extracted from the whole chaotic region from a purely random process (see Figure 5.23 for details).



(a) (b)

Figure 5.24: (a) Histogram showing the distribution of eigenvalues $\rho(\lambda)$ for the Feigenbaum associated to the logistic map ($\mu = 4$) in orange (time series of $N = 2000$), versus the one associated to an iid time series of the same size in purple. To help the eye distinguish both distributions, a smoothing has been applied. (b) Ensemble distribution of the Hellinger distance $H(\mu = 4, \text{iid})$ between $\rho(\lambda)$ for $\mu = 4$ and an iid process, for a total of 100 realisations. The mean of the ensemble is $\langle H(\mu = 4, \text{iid})\rangle = 0.13 \pm 0.03$

5.4.3.2 Distribution of eigenvalues

To round off our analysis, we now compare the distribution of eigenvalues in the chaotic case to the one obtained for random iid time series of the same size. We start with $\mu = 4$. We extract a time series of size $N = 2000$ for each process, compute the list of eigenvalues and display their frequency $\rho(\lambda)$ in a histogram. These are shown in the left panel of Figure 5.24. We observe that the distribution is somewhat different for specific ranges. To quantify 'how different' they are, we compute the Hellinger distance, defined as

$$H(p, q) = \sqrt{1 - \sum_x \sqrt{p(x) \cdot q(x)}},$$

where $p(x)$ and $q(x)$ are two sample distributions. After an ensemble average over 100 realisations, the average Hellinger distance between $\mu = 4$ and iid is $H(\mu = 4, \text{iid}) = 0.13 \pm 0.03$ (see the right panel of Figure 5.24 for the ensemble distribution of Hellinger distances).

Finally, we explore the distance for $\mu \in [\mu_\infty, 4]$. A scatter plot of $H(\mu, \text{iid})$ vs $\text{LE}(\mu)$,

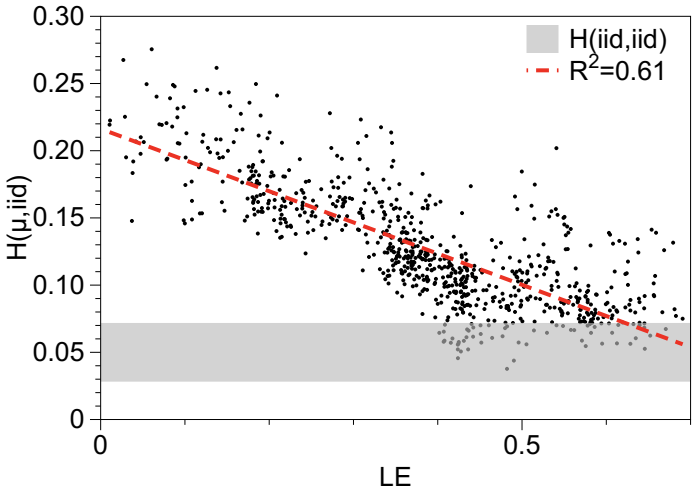


Figure 5.25: Scatter plot of the Hellinger distance $H(\mu, \text{iid})$ between the eigenvalue distribution of the Feigenbaum graph $F(\mu, 2000)$ and the one associated to a random iid time series of the same size as a function of the Lyapunov exponent $\text{LE}(\mu)$, for those values $\mu \in [\mu_\infty, 4]$ for which the Lyapunov exponent is positive (sampling $\Delta\mu = 5 \cdot 10^{-4}$). For comparison, $H(\text{iid}, \text{iid})$ is also shown (the gray area denotes $\langle H(\text{iid}, \text{iid}) \rangle \pm std = 0.05 \pm 0.02$). This area denotes the range for which distributions cannot be distinguished. A clear negative correlation between the Hellinger distance to iid and the Lyapunov exponent emerges.

for those values for which the Lyapunov exponent is positive is shown in Figure 5.25. Unexpectedly, a clear negative correlation emerges between $H(\mu, \text{iid})$ and $\text{LE}(\mu)$. The best linear fit is $H(\mu, \text{iid}) \approx 0.21627 - 0.23208\text{LE}(\mu)$. While a sound theoretical justification for this negative correlation is left for future work, heuristically one can say that the larger the Lyapunov exponent, the more chaotic the time series is and thus the less easy is to distinguish the spectrum of the associated Feigenbaum graph from the one generated from a random series.

5.5 Discussion

Horizontal Visibility Graphs (HVGs) have been widely used as a method to map a time series into a graph representation, with the aim of performing graph-based time series analysis and time series classification. Some of the most popular graph properties used for statistical learning include the degree distribution, entropy of the degree distribution,

and higher order statistics of the degree sequence, and this choice is theoretically justified on the basis that HVGs have been recently shown to be unigraphs [72]. Indeed, several theoretical works have considered the properties of the degree sequence and associated statistics in a number of contexts [4, 36, 84], and in particular a relation between the degree sequence block entropy and the Lyapunov exponent of the associated chaotic map was found [95]. Similarly, the Graph Index Complexity (GIC) (which is a rescaled version of the maximal eigenvalue of the HVG's adjacency matrix) [79] has been used as a network quantifier of the associated time series 'complexity' in several applications [30, 80–83]. Despite this use, and at odds with the theoretical support of the degree sequence, there is currently a shortage of theoretical analysis of the spectral properties of HVGs, as most works essentially deal with applications of GIC for real-world time series classification. This lack of theoretical analysis was a first motivation to study the spectral properties of HVG.

Here we make the first step to partially fill this gap by addressing the spectral properties of HVGs associated to certain classes of periodic and chaotic time series. For convenience, we focus on the archetypal logistic map as it is a canonical system producing both periodic time series of different periods and chaotic time series with different degrees of chaoticity (i.e, different Lyapunov exponent) as it undergoes the Feigenbaum scenario, thereby providing an arena to explore the spectral properties of HVGs associated to different types of dynamics in a single study.

We were able to enumerate the visibility graphs below the map's accumulation point in terms of a bi-parametric family of finite Feigenbaum graphs F_n^k, and have explored their spectral properties (in particular, the behaviour of the maximal eigenvalue of the adjacency matrix) as a function of n (related to the period of the series) and k. We found noteworthy patterns, and numerical results were complemented with analytical developments as well as exact results. Other aspects that were investigated include the full spectrum, the determinant, the number of distinct eigenvalues, and the number of spanning trees of the whole family of F_n^k.

A similar (albeit mostly numerical) analysis was then conducted in the region of the map's parameter where trajectories are chaotic. Unexpectedly, we found that the maximal eigenvalue, while being a good discriminator between chaos and noise, is not able to quantify chaoticity, thereby challenging the previously accepted assumption that GIC is a general indicator of the time series complexity. In this sense, we found that the eigenvalue distribution carried more information about time series chaoticity, in particular its Lyapunov exponent.

Along with these results, in this chapter we have also outlined a number of conjectures and open problems which we hope will trigger some attention in the algebraic and spectral graph theory community. Let us discuss some additional open problems. Consider first the regime $\mu < \mu_\infty$ and the associated family of graphs F_n^k. The limiting fractal-like shape of such spectrum makes us wonder whether the full spectrum of F_n^k could be analytically solved by setting up a recursion on n and k and using e.g. the resolvent formalism subsequently. One could also aim to improve the bounds on $\lambda_{\max}$ for F_n^k, for instance resorting to a isospectral graph reduction scheme [97]. Since F_n^k has a recursive set of left-right symmetries, one could also exploit this hierarchy of automorphisms via the theory of equitable decompositions [98]. A particularly intriguing case is $\mu = \mu_\infty$ (the accumulation point). In such limit, the Feigenbaum graph F_∞^∞ is no longer locally finite. To be able to relate its spectral properties to its dynamical counterpart (i.e., the multifractal properties of the logistic map at the edge of chaos $\mu = \mu_\infty$) also constitutes an intriguing problem [99]. For the region in $\mu > \mu_\infty$ where the map is chaotic, the connection between the spectral properties of the associated graphs and the dynamical properties of this unimodal map has only been explored here numerically, and therefore stating these connections (or the lack thereof) rigorously remain as an open problem. For instance, it would be interesting to rigorously connect the distance of the eigenvalue distribution of the Feigenbaum graph at a certain μ to the i.i.d. case with the Lyapunov exponent of the map at that precise value. In more general terms, the relation between the spectral properties of HVGs and the dynamics of the associated map is of course a

broad open problem of interest.

Chapter 6

Community structure in functional brain networks

6.1 Introduction

Investigating the structure and dynamics of neuronal networks is crucial for understanding the human brain, and the nascent field of "network neuroscience" has yielded fascinating insights into a diverse variety of neurological phenomena [100, 101]. Recent advances in imaging technology have made it possible to perform increasingly detailed investigations of brain structure and dynamics, and it is now possible to map anatomical regions and their interconnecting pathways at near-millimetre resolution. This yields large-scale networks with which to describe the brain's structural connectivity (i.e., the human connectome) [102, 103]. These structural connections govern large-scale neuronal dynamics, which can be captured as patterns of functional connectivity in "functional brain networks" [104–107]. Such functional networks are usually built by estimating coordination or other interdependencies in the neuronal activity of brain regions.

One can measure functional brain networks using various approaches, such as the measurement of blood oxygen level dependent (BOLD) signals gathered via functional mag-

netic resonance image (fMRI) scans or using other modalities [100, 101, 103, 106]. Such studies have yielded many fascinating insights for various disorders and diseases, including Alzheimer's disease [108], autism [109], schizophrenia [110–112], and others [113]. In this chapter, we examine the effects of two antipsychotics (Aripiprazole and Sulpiride) on the architecture of brain functional networks of both controls (who do not have a schizophrenia diagnosis) and patients who have been diagnosed with schizophrenia.

Schizophrenia is often characterized by abnormal and inconsistent social behaviour, along with failure to differentiate between thoughts and reality. Methods for diagnosing schizophrenia have been somewhat controversial [114], and scientists and doctors seek to understand and develop effective diagnoses and treatment (in the form of therapy and drugs) [115]. Sulpiride, a "first-generation antipsychotic" (FGA) and hence a "typical" antipsychotic, works as a selective dopamine agonist and is used for the treatment of schizophrenia [116]. The "atypical" (and thus "second-generation antipsychotic" (SGA)) drug Aripiprazole, which acts as a partial dopamine agonist, is also used to treat schizophrenia [117, 118]. FGAs are cost-effective and have been demonstrated to effectively alleviate positive symptoms, but they carry a risk of extrapyramidal effects (including dystonia, parkinsonism, and tremor). SGAs have the desirable property of avoiding extrapyramidal effects, but they often come with metabolic side-effects and are far more costly. Studies are not conclusive as to which drug type is most effective, and identifying the best course of treatment is a complex issue that varies substantially and must be tailored carefully for each individual [119, 120]. The effectiveness of Sulpiride and Aripiprazole has been reported widely in the literature, and their use for treatment has been approved in many countries [121, 122] (though the United States, Canada, and Australia are notable exceptions). The biological mechanisms of Aripiprazole and Sulpiride are well-understood, but their effects at the functional level of the brain are not. This motivates our goal to explore the effects of these drugs on the architecture of functional brain networks.

It has been hypothesized that schizophrenia is related to abnormalities in the connec-

tivity between components of functional brain networks [110]. An important property of a functional brain network that appears to be abnormal in patients diagnosed with schizophrenia, is community structure [101, 123]. Loosely speaking, a community is a set of nodes in a network that are connected densely to each other but connected sparsely to other parts of a network [124, 125]. Community structure in a network is one type of mesoscale organization, and both community structure and other mesoscale organizations (e.g., core-periphery structure [126]) are important in a variety of contexts in functional brain networks [101]. Although the effects that antipsychotics have on fMRI data have been examined previously [127], few studies have considered the effects of antipsychotics on functional brain networks [112, 128]. In our exploration of such effects, we focus on community structure of functional brain networks and how it is affected by different antipsychotics.

Our research is based on two working hypothesis. The first one is that community structure is a relevant mesoscale structure that may be informative for diagnosing a particular disease. To examine community structure in individuals without schizophrenia (i.e., "controls") versus individuals with schizophrenia ("patients") under the effects of different drugs, we employ several characterizations of graph similarity. We consider both basic features (such as the number of common edges) and more sophisticated ones (such as how community structure changes across different scales of a network [129]). This suite of techniques allows us to build a set of distance matrices between subjects, and we apply unsupervised clustering algorithms to these matrices to try to identify discernible groups of subjects. Our second working hypothesis is that the effects of different antipsychotics leave a measurable fingerprint on a network's community structure. To evaluate this hypothesis, we focus on studying the effects of each drug within a given group (intra-subject comparisons), and we also compare groups who have been given the same drug (inter-subject comparisons). We thereby investigate both the difference between controls and patients and the effects that each of the drugs have on the functional brain networks of each group of subjects.

This chapter proceeds as follows. In Section 6.2, we briefly discuss the employed data set and some relevant previous studies, including a contrast with a recent paper [128]. In Section 6.3, we detail the protocol and the methods that we use to make comparisons between groups of subjects. In Section 6.4, we present our results. Finally, in Section 6.5, we discuss the implications of our findings. We include additional details and technical results in a trio of appendices. We discuss an urn-type model to assess the statistical significance of our results in Section 6.4.3.1.

6.2 Data and Previous Studies

We study a data set, which came from Bristol Myers Squibb (BMS)[1] and which we call the "BMS data set", that consists of measurements of 15 human subjects ("controls") who were deemed to not have schizophrenia and 12 human subjects ("patients") who were diagnosed previously with schizophrenia. All participants were pre-treated with Domperidone on all three days to reduce side effects. Over 3 sessions, which were 1 to 2 weeks apart, each of the 27 subjects was given one of three different drug treatments:

1. ("Placebo") Oral placebo, 180 and 90 minutes before scanning;

2. ("Sulpiride") Oral placebo, 180 minutes before scanning; and then oral Sulpiride (400 mg), 90 minutes before scanning;

3. ("Aripiprazole") Oral Aripiprazole (15 mg), 180 minutes before scanning; and then oral placebo, 90 minutes before scanning.

All participants and investigators were blind to the drug condition. All participants were provided with a detailed Patient Information Sheet (PIS) that explained the nature of the pharmacological experiment (comparison of single doses of the two drugs being used for the treatment of schizophrenia with placebo pill) and the double-dummy design.

At each session, after being given one of the drug treatments, each individual was placed

[1]Data collection was supported by a grant from Bristol Myers Squibb to Robert Kerwin at King's College London

in an fMRI scanner to measure blood flow, at resting state, in the brain. The fMRI scanner captures a single image once every 2 seconds. The scans lasted 17 minutes and 4 seconds, so each BOLD time series has 512 time points. The data are parcellated into 298 regions of interest (RoIs), and each region corresponds to a node in a functional brain network. We used an anatomically-driven parcellation scheme and methodology, as described in [130], to partition the data for each subject into 325 contiguous regions, which were as uniform as possible. However, 27 regions did not have high-quality fMRI time series for one or more individuals and were later removed from all subjects, leading to a total of 298 homogeneously-sized regions. Each region has a corresponding time series that represents an average level of activity in that region. We remove 4 controls (2, 8, 10, and 14) and 3 patients (3, 5, and 11) from our calculations due to missing data and/or problems due to head motion. We thus examine a total of 20 subjects: 11 controls and 9 patients. (However, we use the original numerical labels for the subjects.) See [131–133] for discussions of issues with head motion, and see [134, 135] for discussions of preprocessing of fMRI data to correct for head motion.

There have been three other studies [110, 128, 136] that employed this particular data set. References [110, 136], which were published a few years ago, focused on the task of distinguishing controls from patients who had been diagnosed with schizophrenia; the aim was to find effective biomarkers for schizophrenia. Using a parcellation with 90 RoIs, Ref. [110] reported that the patients have "less strongly connected" brain networks (in the sense of a lower mean pairwise wavelet coherence between regions) and "more diverse" profiles (in the sense of larger mean variances in a wavelet coherence between a given region and the others) of brain functional connectivity than the controls. They also calculated that brain networks in the schizophrenia group have a greater robustness to uniform-at-random removal of nodes, in the sense that the number of nodes in the largest connected component (LCC) decays more slowly as a function of the number of removed nodes. Reference [136] built functional networks via "spatial pairwise clustering" (a novel approach that they introduced) of individual voxels (thereby foregoing the need to

choose a parcellation) and combining spatially proximate voxels into nodes (similar to the method in Section 4.3.2). In their computations, they observed weaker inter-nodal correlations in patients than in controls. Finally, using a very similar parcellation to the one that we employ but with different techniques from network analysis, a very recent work [128] studied the effects of the drugs on (1) the networks of the subjects and (2) their cognitive abilities. Their results suggest that (1) Aripiprazole has a major effect on the networks of controls and that (2) both drugs make it harder to distinguish controls and patients. This study also found that Aripiprazole diminished the performance of controls at a working-memory task.

6.3 Methods and Preliminary Computations

We illustrate our analysis pipeline with a schematic in Figure 6.1. In Subsections Section 6.3.1 and Section 6.3.2, we briefly describe how to build a functional network from fMRI time series using wavelet correlations and thresholding techniques (step 1 in Figure 6.1). In Section 6.3.3, we discuss our preliminary computations on our collection of networks. In Section 6.3.4 and Section 6.3.5, we discuss how to define two distance functions to examine dissimilarities of functional networks (step 2 in Figure 6.1) and how to apply hierarchical clustering to cluster similar subjects (i.e., similar functional networks) according to step 3 in Figure 6.1).

6.3.1 Building the Networks

Wavelet-based correlations allow one to examine functional similarities between brain regions based on activity in a specified frequency interval (a wavelet "scale"). We use the maximal-overlap discrete wavelet transform [137] to decompose each regional mean fMRI time series (see step 1 in Figure 6.1). Examining wavelets is useful for studying resting-state fMRI data, and functional connectivity between regions is typically largest at certain frequency bands (below 0.1 Hz) [138]. Let g_i denote the time series of node (i.e., RoI) i (where $i \in \{1, 2, \ldots, 298\}$), and let $V_s(g_i)$ denote the vector of scale-s wavelet

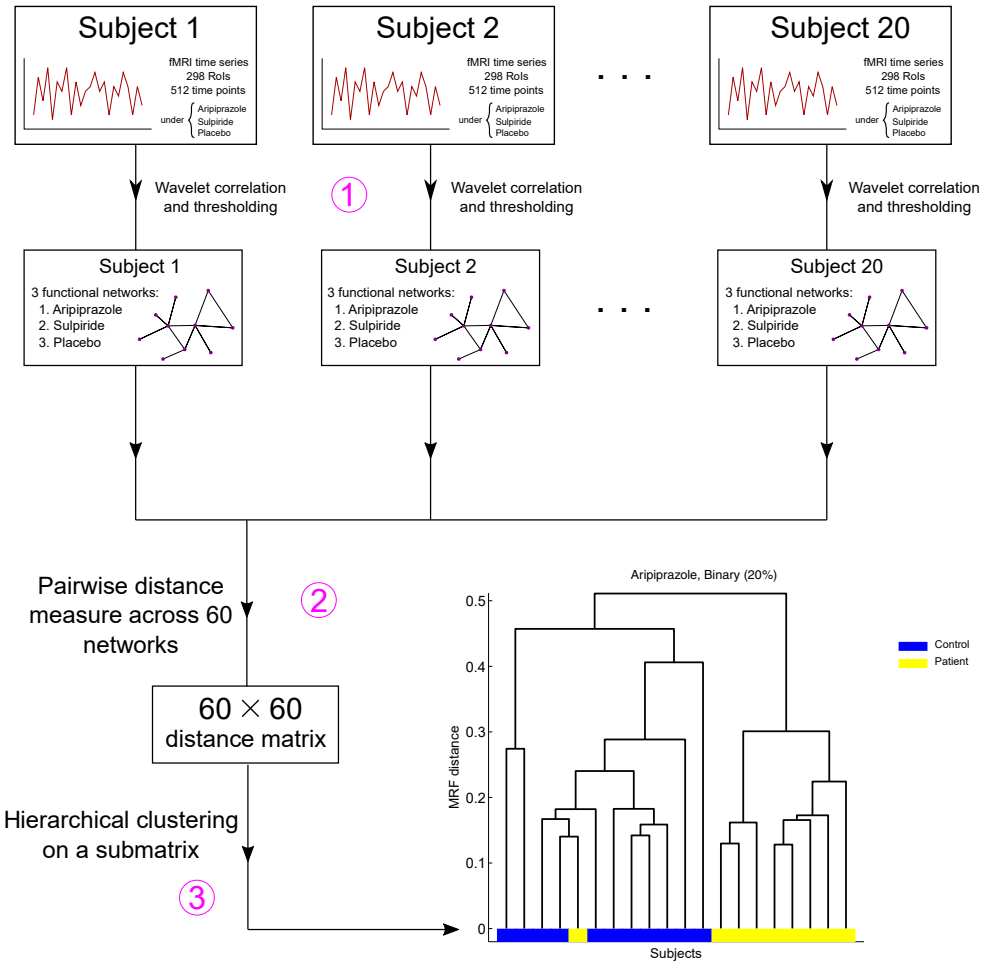


Figure 6.1: Protocol to obtain a dendrogram that conveys hierarchical clustering of a set of subjects. There are 20 subjects, and there are 3 different drug treatments for each subject. This yields 60 networks, and we compute a distance between each pair of networks. This yields a 60×60 distance matrix. (As we discuss in Section 6.3.4, we construct such a matrix for two different notions of distance.) We do hierarchical clustering using various submatrices of each distance matrix, where the submatrix that we use depends on our particular comparison from Figure 6.6. We explain the vertical axis (which uses a particular choice of distance) in the dendrogram in Section 6.3.4.2. In the example dendrogram in this schematic, we consider unweighted networks that include the strongest 20% of the edges (see Section 6.3.1).

coefficients of g_i. At scale s, the connection strength between two nodes, i and j, in a functional network is given by the wavelet correlation

$$F_{ij} = \frac{\sum_k V_{s,k}(g_i) V_{s,k}(g_j)}{\sqrt{(\sum_k (V_{s,k}(g_i))^2 (\sum_k (V_{s,k}(g_j))^2}} \in [-1, 1]\,. \tag{6.1}$$

We compute values of F_{ij} for scales $s = 1, 2, 3, 4$; and we then choose to work with the most informative scale (see Section 6.4).

There are $N = 298$ RoIs for each subject, so we extract functional networks with $N = 298$ nodes. This yields a similarity matrix $\mathbf{F}$ whose elements are given by Equation (6.1). To avoid negative weights[2], we transform $\mathbf{F}$ into a weighted adjacency matrix $\mathbf{W}$ by taking $W_{ij} := (F_{ij} + 1)/2 \in [0, 1]$. The associated network is fully connected by construction, and there are two customary ways to prune edges. These are (1) thresholding the networks by keeping a fixed fraction τ of the strongest weights (assigning the remaining edges a weight of 0 and producing thresholded weighted networks) and (2) first performing the previous step and then subsequently setting the remaining edges to have a weight of 1, thereby producing thresholded binary networks. In both cases, the resulting thresholded networks have $E = N(N-1)\tau/2$ edges. Of course, one can also simply keep all edges and examine the original fully connected, weighted networks. We initially examine the original networks and both the weighted and binary thresholded networks. Based on some preliminary calculations, we then decide which of these networks to examine further.

6.3.2 Choosing a Scale and Thresholding Parameter

To construct the functional networks, we choose a wavelet scale s and then consider thresholding the networks (with an associated threshold value). Previous work has noted differences in both "connectivity" (i.e., the mean edge weight of a network) and mean local clustering coefficient between controls and patients with schizophrenia [110, 140, 141]. The observed differences were more statistically significant at lower frequencies, and they were particular evident at scale 2. This is consistent with previous research on resting-state fMRI [142]. To make an educated choice of scale, Ref. [110] calculated

[2]There are also other ways to transform $\mathbf{F}$ into a weighted adjacency matrix $\mathbf{W}$. For example, one can take the absolute value of the similarity values, though it is then impossible to distinguish negative wavelet similarities from positive ones. The weakness of our chosen approach is that we transform initially strongly negative weights into weights that are near 0, and they then tend to be removed if one subsequently prunes a network by keeping only the most strongly weighted edges of $\mathbf{W}$. Recently, [139] examined the significance of such negative wavelet similarities.

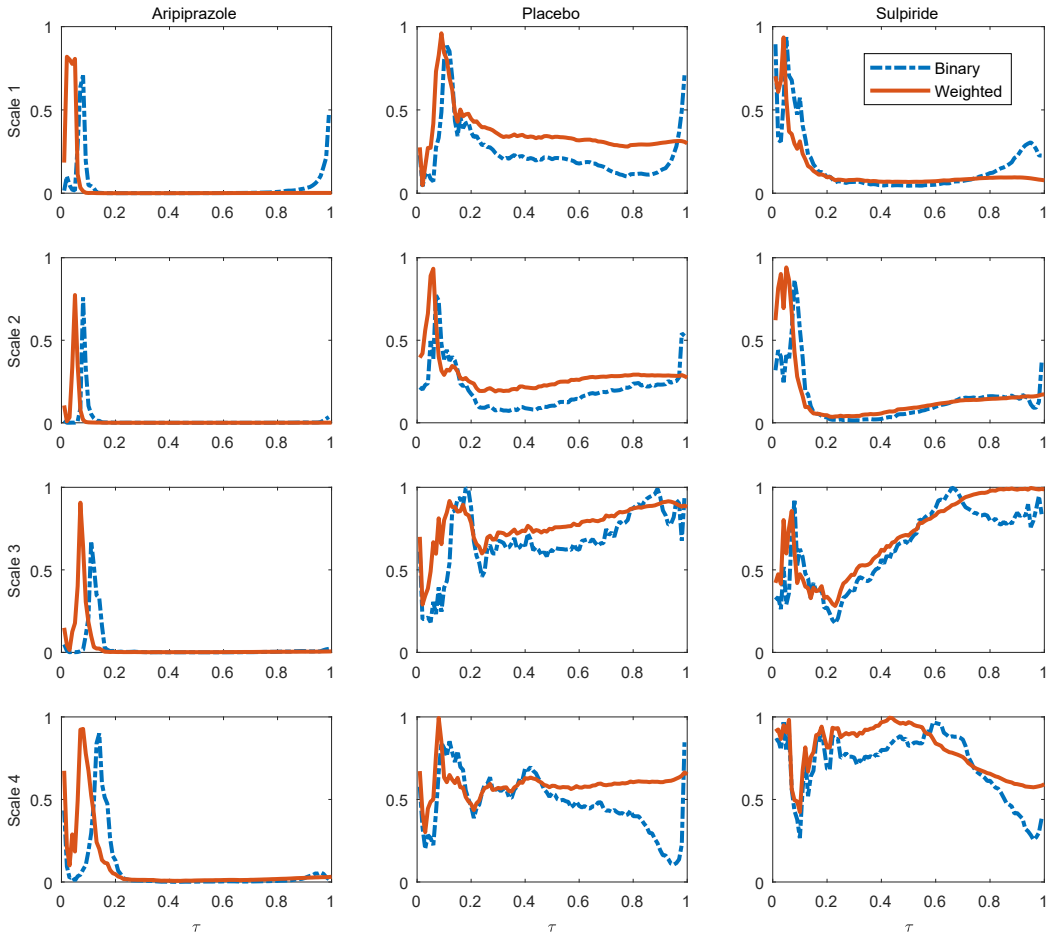


Figure 6.2: The p-values associated with t-tests on the mean local clustering coefficient (between patients and controls) for weighted networks (solid orange curves) and binary networks (blue dashed curves) for different values of the thresholding parameter τ. Wavelet scale 2 produces the smallest p-values. We also observe differences in the curves associated with the three drug treatments and that the p-values associated with the binary networks are consistently smaller than those for the weighted ones.

the mean value of F_{ij} over controls and patients for each scale, performed a t-test, and selected the scale with the smallest p-value. We follow a similar procedure, but we also threshold the networks for both binary and weighted versions using a thresholding parameter τ, such that we keep a fraction τ of the strongest edges (i.e., those with the largest weights).[3] (For example, if $\tau = 0.4$, we keep the strongest 40% of the edges.) For each of the three drug treatments and for each of the scales 1, 2, 3, and 4, we then perform a t-test on the mean local clustering coefficients of controls and patients. In Figure 6.2,

[3]We consider values of τ in increments of 0.01.

we show all 12 plots and the p-values associated with the t-tests. Based on these results, we make two decisions. First, from now on, we use scale 2 (which corresponds to the frequency band 0.060–0.125 Hz), because it has the smallest p-values (in agreement with previous work [110]). For very small values of τ, we observe spikes in the p-values that likely arise from the networks breaking up into many components. Second, because our results on binary networks have smaller p-values than the corresponding ones for weighted networks, we focus our subsequent calculations on thresholded binary networks (except for our calculations of connectivity). The controls tend to have much larger edge weights than the patients, so our comparisons between patients and controls are more directly parallel if we use binary networks, as many network quantities are affected in nontrivial ways by edge weights. From now on, we fix $\tau = 0.2$. (We repeat our calculations for several values of $\tau \in [0.2, 0.4]$, and we obtain qualitatively similar results.)

Half of our networks (30 out of 60) have more than one component when $\tau = 0.2$. This can be problematic for some types of computations, such as those that involve path lengths. In practice, however, this issue has not been problematic for us, and the largest connected component of each network has almost the maximum of 298 nodes, with the exception of Control 5 on Aripiprazole, whose largest connected component has 268 nodes. In Section 6.6, we show the number and sizes (i.e., number of nodes) of the components in each of our networks.

6.3.3 Connectivity and Mean Local Clustering Coefficient

We now do some preliminary calculations. Previous research using thresholded, binary networks has highlighted significant differences in "connectivity" (defined, for an individual subject, as the mean edge weight $\langle W_{ij} \rangle$ of a network) and mean local clustering coefficients of networks from control subjects versus those from patients diagnosed with schizophrenia [110]. In our case, by construction, connectivity corresponds (up to a scaling and a shift) to the mean wavelet correlation. For weighted networks, we compute

the weighted local clustering coefficient[143]

$$c_i = \frac{1}{k_i(k_i-1)} \sum_{j,k} (W_{ij} W_{ik} W_{jk})^{1/3} \quad \text{for } k_i \geq 2\,, \tag{6.2}$$

where k_i is the degree of node i and $c_i = 0$ for $k_i \in \{0, 1\}$. Equation (6.2) reduces to the usual local clustering coefficient for the special case of binary networks.

For connectivity, we calculate $\langle W_{ij} \rangle$ for each subject, and we then calculate the means for both controls and patients. We follow the same process for the local clustering coefficient. In our preliminary analysis, we explore how these basic quantities differ for different drug treatments. Specifically, we calculate connectivity using the non-thresholded weighted versions of the networks and mean local clustering coefficient using the thresholded binary networks. We show our results in Figure 6.3, where for each case we plot the mean and standard deviation across subjects. For each drug treatment, we also perform a two-sample t-test on the values of connectivity and mean local clustering coefficients for controls and patients, and we extract a p-value. We observe small differences in connectivity and mean local clustering coefficients between controls and patients; this difference is smaller than what was reported previously with these data using other approaches [110]. We also observe that Aripiprazole has a small effect on the connectivity and mean local clustering coefficients of controls but no significant effect on patients, in agreement with other recent work [128]. Sulpiride appears to have little effect on either group, though we observe a larger difference between controls and patients for mean local clustering coefficient than we do for connectivity. We obtain a p-value of $p \approx 0.0326$ for mean local clustering coefficient and a p-value of $p \approx 0.1680$ for connectivity. We show the connectivity for all subjects under placebo in Figure 6.4, and we note that Patient 8 has a very large value of connectivity. However, given the sizes of the error bars, we cannot reject the hypotheses that the connectivity and/or mean local clustering coefficients are indistinguishable in the different situations. This suggests that either (1) this data set is not large enough for these measures to detect robust differences, and/or that (2) these simple network diagnostics may not give clear information about whether

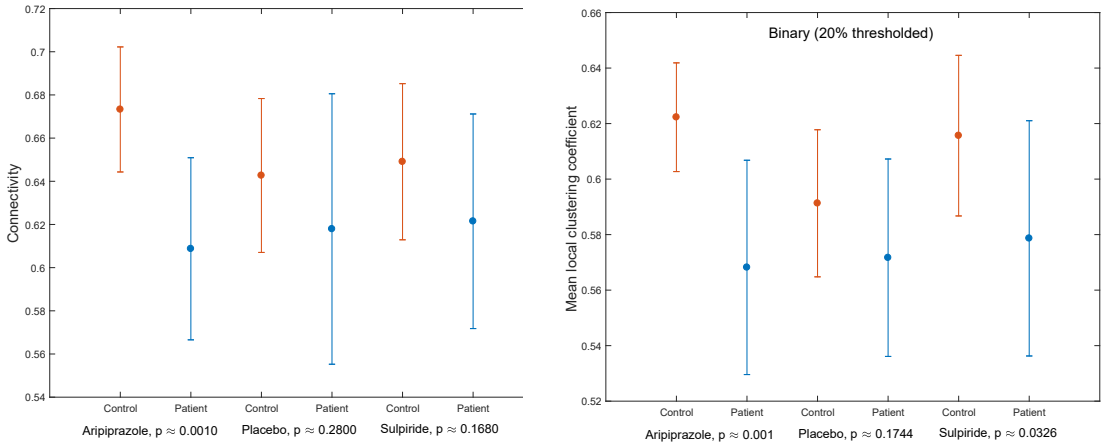


Figure 6.3: Means and standard deviations of (left) connectivity for non-thresholded weighted networks and (right) mean local clustering coefficients for binary networks thresholded to 20% of the strongest edges. The results are similar in each case, although we observe for Sulpiride that the controls and patients have different p-values for the two-sample t-test.

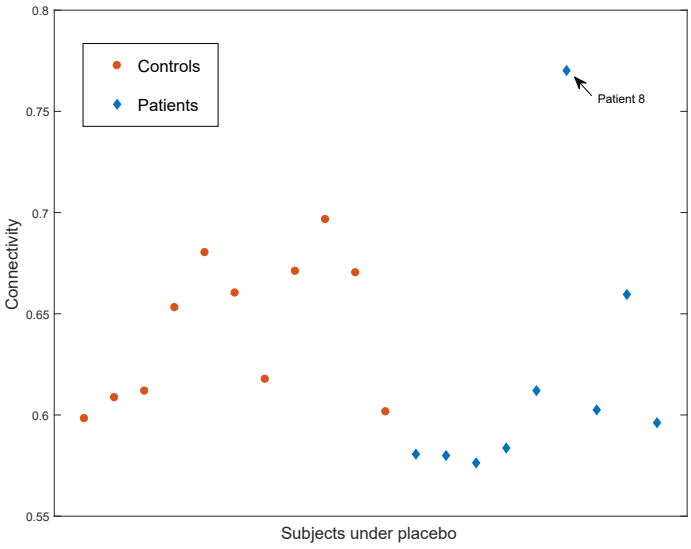


Figure 6.4: Connectivity of each subject under placebo. We observe that Patient 8 has an abnormally large value of connectivity. (Recall that we removed Patients 3 and 5 from consideration because of missing data and problems with head motion, but we use the original numerical labelling of the subjects.)

the drugs have any discernible effects on the architecture of functional brain networks.

Given the inconclusiveness of these results, we need to do a more sophisticated analysis.

6.3.4 Distance Measures

As we mentioned in Section 6.1, we aim to classify similar functional brain networks using unsupervised clustering of subjects. A subject is associated with a functional network. To classify these networks in a systematic way, we define a pairwise distance function between graphs, and we then use this function to compute a distance matrix for a set of subjects. (See step 2 in Figure 6.1.) We consider distance functions based on two rather different aspects of networks.

6.3.4.1 Hadamard-like distance

One can construct a simple similarity measure between binary networks $\mathbf{A}$ and $\mathbf{B}$ that both have the same number of edges by computing the Hadamard product of the matrices and then summing the entries $A_{ij}B_{ij}$ of the resulting matrix. For binary networks, this sum ($\sum_{i>j} A_{ij}B_{ij}$) is the number of common edges in the networks. Because it is common to threshold functional networks so that one retains only a specified, fixed fraction of edges, we can use this similarity measure to compare adjacency matrices that we extract from thresholded functional networks. We define the function

$$d_1(\mathbf{A}, \mathbf{B}) = 1 - \frac{1}{E} \sum_{i>j} A_{ij}B_{ij} \in [0, 1]\,, \tag{6.3}$$

which is well-defined when $\mathbf{A}$ and $\mathbf{B}$ have the same number E of edges, and is indeed a metric which we will prove in Theorem 6.3.1. We can then construct a distance matrix $\mathbf{D}^1$, whose elements $\mathbf{D}^1_{\alpha\beta}$ measure the distance between the functional networks of subjects α and β. Using $\mathbf{D}^1$ has the advantage of being computationally efficient and based on a mathematically sound metric, although d_1 is a rather simplistic measure (two networks are more distant from each other when they have fewer common edges) and we do not expect it to capture certain details (e.g., community structure) of the networks. We will now state a theorem and a proof, which as a trivial corollary will give us a proof that Equation (6.3) is a metric.

Theorem 6.3.1. *Let $\mathcal{S}_n(E)$ be the set of $n \times n$ square matrices with entries of 0 or*

1 *(i.e., "binary matrices"), where the number* E *of* 1 *entries satisfies* $E < n^2$*. Let* $\mathbf{A}, \mathbf{B} \in \mathcal{S}_n(E)$ *be two arbitrary elements of the set. Consider the function* d_1 *defined by*

$$d_1 : \mathcal{S}_n(E) \times \mathcal{S}_n(E) \to [0, 1], \quad d_1(A, B) = 1 - \frac{1}{E} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij} . \tag{6.4}$$

The function d_1 *is a metric.*

The function in Equation (6.4) for d_1 is slightly more general than the function defined in Equation (6.3), as here we are not assuming that (1) $\mathbf{A}$ and $\mathbf{B}$ are symmetric or that (2) there are no 1 entries in the diagonal (E is the number of 1 entries). In Equation (6.3) we imposed some restrictions on E that are not present here; we used E to denote the number of edges in an associated network, so for Equation (6.3) (which is designed to deal with unweighted, undirected adjacency matrices with no self-loops), one needs either to restrict to the case in which there are no 1 entries in the main diagonal and then do the relabelling $E \to E/2$ or to relabel the summation indices with $\sum_{i=1}^{n} \sum_{j=1}^{n} \to \sum_{i>j}$. The proof that the function defined in Equation (6.3) is a metric follows as a trivial corollary of Theorem 6.3.1.

Proof. To prove that d_1 is a metric, we need to prove four properties: nonnegativity, identity of indiscernibles, symmetry, and the triangle inequality. The first three properties are satisfied trivially:

1. Nonnegativity: By construction, $\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij} \leq E$, so $d_1(\mathbf{A}, \mathbf{B}) \geq 0$.

2. Identity of indiscernibles: $d_1(\mathbf{A}, \mathbf{B}) = 0 \Leftrightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij} = E$. However, by definition, the matrices are binary and have E entries with the value 1, so $\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij} = E \Leftrightarrow \mathbf{A} = \mathbf{B}$.

3. Symmetry: This arises trivially from the commutative property of the scalar product: $A_{ij} B_{ij} = B_{ij} A_{ij}$.

To prove the fourth property (the triangle inequality), we need to show that

$$\text{for all } \mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{S}_n(E)\,, \quad d_1(\mathbf{A}, \mathbf{B}) + d_1(\mathbf{B}, \mathbf{C}) \geq d_1(\mathbf{A}, \mathbf{C})\,. \tag{6.5}$$

This part is more subtle, and we need to break the proof into a couple of steps. We start by defining a *matrix δ-perturbation*.

Definition 6.3.2 (Matrix δ-perturbation)**.** Let $\mathbf{A} \in \mathcal{S}_n(E)$, and let δ be a positive integer such that $0 < \delta < E$. The matrix $\tilde{\mathbf{A}}^{(\delta)}$ is a δ-perturbation of $\mathbf{A}$ if $\tilde{\mathbf{A}}^{(\delta)}$ is constructed by taking $\mathbf{A}$ and changing the position of δ of the 1 entries.

To illustrate this definition, we show an example of a matrix and a 1-perturbation of that matrix in $S_3(3)$:

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\,, \quad \tilde{\mathbf{Z}}^{(1)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\,. \tag{6.6}$$

It is clearly the case that $\tilde{\mathbf{A}}^{(\delta)} \in \mathcal{S}_n(E)$. It is also true that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\tilde{A}_{ij}^{(\delta)} = E - \delta \;\Rightarrow\; d_1(\tilde{\mathbf{A}}^{(\delta)}, \mathbf{A}) = \delta/E\,.$$

Starting from an arbitrary element of $\mathcal{S}_n(E)$, one can reach any other element by applying an appropriate δ-perturbation. Therefore, equipped with the δ-perturbation, $\mathcal{S}_n(E)$ is a unary system. This property is important for guaranteeing completeness.

To prove Equation (6.5), it is equivalent to prove that

$$\text{for all } \mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{S}_n(E)\,, \quad \mathcal{X} := \sum_{i=1}^{n}\sum_{j=1}^{n}(A_{ij}B_{ij} + B_{ij}C_{ij} - A_{ij}C_{ij}) \leq E\,.$$

We are ready to prove this latter inequality. We start with a degenerate case. Consider an arbitrary $\mathbf{A} \in \mathcal{S}_n(E)$ and set $\mathbf{A} = \mathbf{B} = \mathbf{C}$; in this case, $\mathcal{X} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}A_{ij} = E \leq E$.

To generate all possible triples $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$, without loss of generality, we now consider an

arbitrary (but fixed) $\mathbf{A} \in \mathcal{S}_n(E)$; and we use δ-perturbations to generate all instances of $\mathbf{B}$ and $\mathbf{C}$. That is,

$$\mathbf{B} := \tilde{\mathbf{A}}^{(\delta_b)},\ \mathbf{C} := \tilde{\mathbf{A}}^{(\delta_c)},\ \delta_b, \delta_c \geq 0\,.$$

All possible triples can be expressed in this form.

Let's evaluate $\mathcal{X}$. The first term is

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\tilde{A}_{ij}^{(\delta_b)} = E - \delta_b\,;$$

the second term is

$$\sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}C_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{A}_{ij}^{(\delta_b)}\tilde{A}_{ij}^{(\delta_c)}\,;$$

and the third term is

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}C_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\tilde{A}_{ij}^{(\delta_c)} = E - \delta_c\,.$$

We need to separately consider the cases in which a pair of matrices experience the same perturbation or different perturbations. In the usual case, $\delta_b \neq \delta_c$ (i.e., the perturbations are different), so there is at least an offset of $|\delta_b - \delta_c|$. Consequently,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{A}_{ij}^{(\delta_b)}\tilde{A}_{ij}^{(\delta_c)} \leq E - |\delta_b - \delta_c|\,. \tag{6.7}$$

If, however, $\delta_b = \delta_c$ (i.e., both δ-perturbations are the same), the right-hand-side of Equation (6.7) is instead given by E.

Altogether, this yields the following:

$$\mathcal{X} \leq E - \delta_b + E - |\delta_b - \delta_c| - E + \delta_c = E + (\delta_c - \delta_b) - |\delta_b - \delta_c|\,.$$

Three possibilities emerge:

1. If $\delta_b = \delta_c$, then $\mathcal{X} \leq E$.

2. If $\delta_b < \delta_c$, then $|\delta_b - \delta_c| = \delta_c - \delta_b$, so $\mathcal{X} \leq E$.

3. If $\delta_b > \delta_c$, then $|\delta_b - \delta_c| = \delta_b - \delta_c$, so $\mathcal{X} \leq E + 2(\delta_c - \delta_b) < E$.

This concludes the proof. □

6.3.4.2 Distance based on community structure

We also use a more sophisticated distance measure, introduced by Onnela et al. [129], that is based on network community structure [124, 125]. It requires using a method of partitioning that assigns each node to a community (i.e., it is a "hard partition"). Here we use modularity maximization [144, 145] and employ the code of Onnela et al. that implements the (locally greedy) Louvain method [146].

Given a network described by its weight matrix $\mathbf{W}$, one can detect communities in it by maximizing modularity, which one does by minimizing the objective function

$$\mathcal{H}(\gamma) = -\sum_{i \neq j} \left(W_{ij} - \gamma \frac{r_i r_j}{2M} \right) \delta(C_i, C_j) \,, \tag{6.8}$$

where γ is a resolution parameter, C_i is the community assignment of node i (and C_j is the community assignment of node j), r_i is the strength (i.e., sum of incident edge weights) of node i, and M is the total edge weight. We consider undirected networks, so we use the Newman-Girvan null-model matrix $\mathbf{P}$ with elements $P_{ij} = r_i r_j/(2M)$ [145, 147]. The quantity $W_{ij} - P_{ij}$ is the "effective weight" of the edge between nodes i and j. For unweighted networks, node strength reduces to degree (i.e., $r_i = k_i$ and $r_j = k_j$), and the total edge weight reduces to the total number of edges (i.e., $M = E$). For each value γ, minimizing the objective function in Equation (6.8) gives a partition of nodes into disjoint communities. The quantity $\mathcal{H}(\gamma)$ also quantifies the (scaled) energy of the system [148]. For illustration, we show a particular partition of a brain functional network into communities in the left panel of Figure 6.5. (See the middle panel of the same figure for the same network embedded in a three-dimensional (3D) physical space, where node locations correspond to the actual physical regions.)

Onnela et al. defined "mesoscopic response functions" (MRFs) for three quantities that

describe, from different perspectives, how a partition of a network changes as a function of γ. In calculating a network's MRF, one varies the parameter γ between $\gamma_{\min}$ (where community detection yields a single community) and $\gamma_{\max}$ (where each node is assigned to its own community). Let n_k denote the number of nodes in community k and define $p_k = n_k/N$ to be the probability of choosing a node uniformly at random from community k. One can then define a partition entropy of the associated community-size distribution as $\mathcal{S}(\gamma) = -\sum_{k=1}^{\eta(\gamma)} p_k \log(p_k)$, where $\eta(\gamma)$ is the number of communities. One then defines the effective energy ($\mathcal{H}_{\text{eff}}$), effective entropy ($\mathcal{S}_{\text{eff}}$), and the effective number of communities (η_{eff}) as follows:

$$\mathcal{H}_{\text{eff}}(\gamma) = \frac{\mathcal{H}(\gamma) - \mathcal{H}(\gamma_{\min})}{\mathcal{H}(\gamma_{\max}) - \mathcal{H}(\gamma_{\min})} = 1 - \frac{\mathcal{H}(\gamma)}{\mathcal{H}(\gamma_{\min})}\,, \tag{6.9}$$

$$\mathcal{S}_{\text{eff}}(\gamma) = \frac{\mathcal{S}(\gamma) - \mathcal{S}(\gamma_{\min})}{\mathcal{S}(\gamma_{\max}) - \mathcal{S}(\gamma_{\min})} = \frac{\mathcal{S}(\gamma)}{\log(N)}\,, \tag{6.10}$$

$$\eta_{\text{eff}}(\gamma) = \frac{\eta(\gamma) - \eta(\gamma_{\min})}{\eta(\gamma_{\max}) - \eta(\gamma_{\min})} = \frac{\eta(\gamma) - 1}{N - 1}\,. \tag{6.11}$$

One uses a parameter ξ that tracks, in a discrete manner (keeping track of when each effective weight changes sign), which edges have an positive effective weight and which have a negative effective weight. By construction, varying γ from $\gamma_{\min}$ to $\gamma_{\max}$ corresponds to varying ξ from 0 to 1. For a detailed discussion, see [129].

To each network, one associates a curve of each of $\mathcal{H}_{\text{eff}}$, $\mathcal{S}_{\text{eff}}$, and η_{eff} (or for any other quantity that one wishes to track [149]) as a function of ξ; these are the MRFs. In the right panel of Figure 6.5, we show example MRFs that we compute from our functional brain networks. We show average MRFs (which we compute as a pointwise mean of the MRFs for the 60 networks), along with the minimum and maximum MRF (which we determine based on ordering the area under the curve of each MRF from largest to smallest), of these networks.

To compare a pair of networks, we compare the differences in their profiles. Consider a

pair of networks, α and β, along with the following three distances:

$$d_{\alpha\beta}^{\mathcal{H}} = \int_0^1 \left| \mathcal{H}_{\text{eff}}^{\alpha}(\xi) - \mathcal{H}_{\text{eff}}^{\beta}(\xi) \right| d\xi \,, \tag{6.12}$$

$$d_{\alpha\beta}^{\mathcal{S}} = \int_0^1 \left| \mathcal{S}_{\text{eff}}^{\alpha}(\xi) - \mathcal{S}_{\text{eff}}^{\beta}(\xi) \right| d\xi \,, \tag{6.13}$$

$$d_{\alpha\beta}^{\eta} = \int_0^1 \left| \eta_{\text{eff}}^{\alpha}(\xi) - \eta_{\text{eff}}^{\beta}(\xi) \right| d\xi \,. \tag{6.14}$$

The trio of distances in Equations (6.12) to (6.14) capture different aspects of community structure. The effective energy ($\mathcal{H}_{\text{eff}}$) is a rescaled version of the objective function $\mathcal{H}$, the effective entropy ($\mathcal{S}_{\text{eff}}$) quantifies the amount of heterogeneity in the sizes of the detected communities, and the effective number of communities (η_{eff}) is a rescaled version (with respect to network size) of the total number of communities. From these distances matrices, we construct a single distance matrix by projecting each 3D coordinate using principal component analysis (first mentioned in Section 3.3.3) and keeping the first component. That is, we construct a distance matrix by calculating a linear combination of the three distance measures:

$$d_{\alpha\beta}^{P} = w_{\mathcal{H}} d_{\alpha\beta}^{\mathcal{H}} + w_{\mathcal{S}} d_{\alpha\beta}^{\mathcal{S}} + w_{\eta} d_{\alpha\beta}^{\eta} \,, \tag{6.15}$$

where the weights w_ℓ (with $\ell \in \{\mathcal{H}, \mathcal{S}, \eta\}$) are the coefficients of the first principal component. There are a total of 60 networks (11 controls and 9 patients, each of which is on 3 different drug treatments). We calculate the matrix composed of $60 \times 59/2$ (the total number of network pairs) rows and 3 columns, where each column corresponds to the vector representation of the upper triangle of one of the distance matrices $\mathbf{D}^{\mathcal{H}}$, $\mathbf{D}^{S}$, and $\mathbf{D}^{\eta}$. We perform a PCA on this matrix to create a distance matrix $\mathbf{D}^{P}$.

The final outcome of the above calculation is a 60×60 distance matrix $\mathbf{D}^{P}$, where each entry measures the distance between networks α and β based on how the community structure of each network varies as a function of the parameter ξ. We henceforth use the

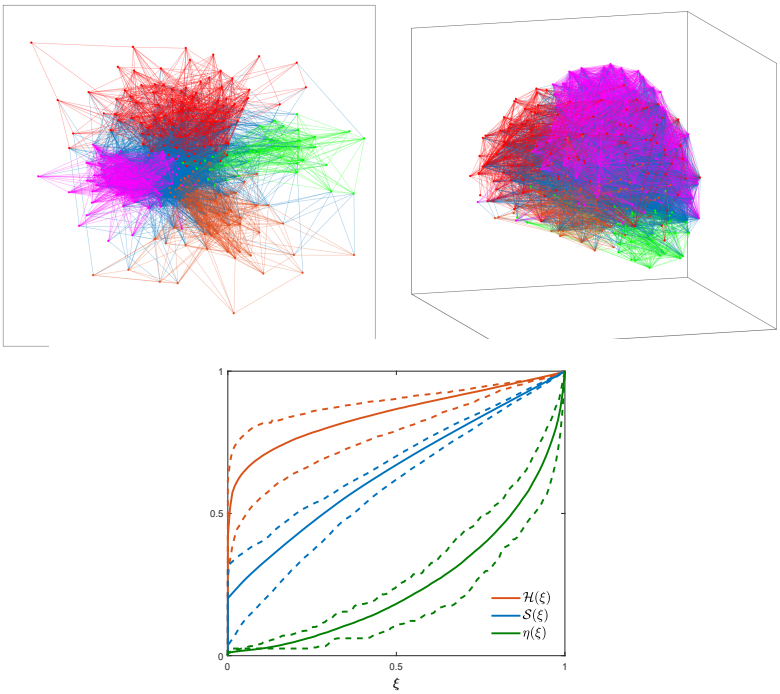


Figure 6.5: An example, which we show both (left) in 2D and (right) in 3D, of a particular network (Control 1 on placebo) partitioned into communities for a specific value of the resolution parameter ($\gamma = 1$, so $\xi \approx 0.02$). (Right) An average MRF, which we determine by taking a pointwise mean of the MRFs of all 60 networks, along with the maximum and minimum curves (based on the area under each MRF curve). For each colour, the upper curve is the maximum, the middle curve is the pointwise mean, and the bottom curve is the minimum. We show $\mathcal{H}_{\text{eff}}(\xi)$, as defined in Eq. Equation (6.9), in orange; we show $\mathcal{S}_{\text{eff}}(\xi)$, as defined in Equation (6.10), in blue; and we show $\eta_{\text{eff}}(\xi)$, as defined in Equation (6.11), in green.

term "MRF distance" for the quantity that we compute in Equation (6.15).

6.3.5 Hierarchical Clustering

Once we have our distance matrix (see Section 6.3.4.2), we take a submatrix of it for each of the comparisons in Figure 6.6. For example, if we are comparing controls and patients under the drug Aripiprazole, we keep only the rows and columns that correspond to this drug, leaving us with a 20×20 distance matrix, where the rows and columns correspond to the 11 controls and 9 patients. We then cluster the new, smaller distance matrix using one of numerous possible methods. For simplicity, we use average linkage clustering to group similar subjects (i.e., similar networks) together and show our results

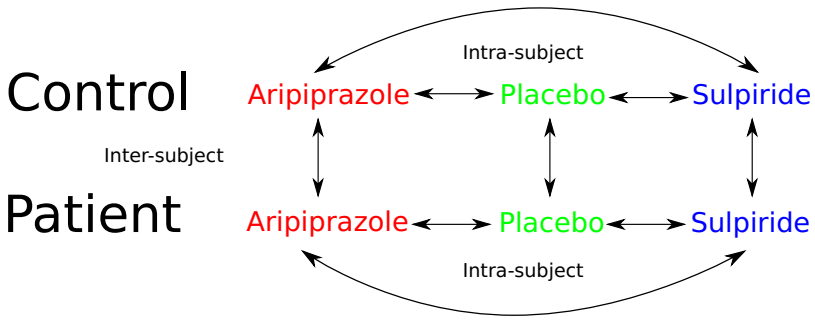


Figure 6.6: Illustration of possible comparisons between the groups of subjects and different drug treatments.

in the form of dendrograms. We then order the leaves of the dendrogram to maximize the sum of the similarities between adjacent leaves by reordering its branches (without further partitioning clusters). We colour the leaves of the dendrograms based on their annotations: patients or controls without drugs, patients or controls on one drug, or patients and controls on the other drug.

6.4 Main Results

As we mentioned in Section 6.1 and depicted in Figure 6.6, we make a total of 9 comparisons, including both inter-subject ones (different groups under the effect of the same drug) and intra-subject ones (the same group under the effect of different drugs). In our ensuing discussions, we present the results of these comparisons.

6.4.1 Inter-subject Comparisons

We do inter-subject comparisons using the procedure that we outlined in Figure 6.1. We start by comparing controls and patients under the effects of the drug Aripiprazole using the simple distance measure $\mathrm{d}_1(\mathbf{A}, \mathbf{B})$ from Equation (6.3). We show the resulting dendrogram in Figure 6.7. We observe some separation between patients and controls.

To do a more sophisticated analysis, we compute a dendrogram on the same data using the MRF distance matrix $\mathbf{D}^P$ (see Section 6.3.4). We show the resulting dendrogram in Figure 6.8. The separation between patients and controls is now better, and we correctly

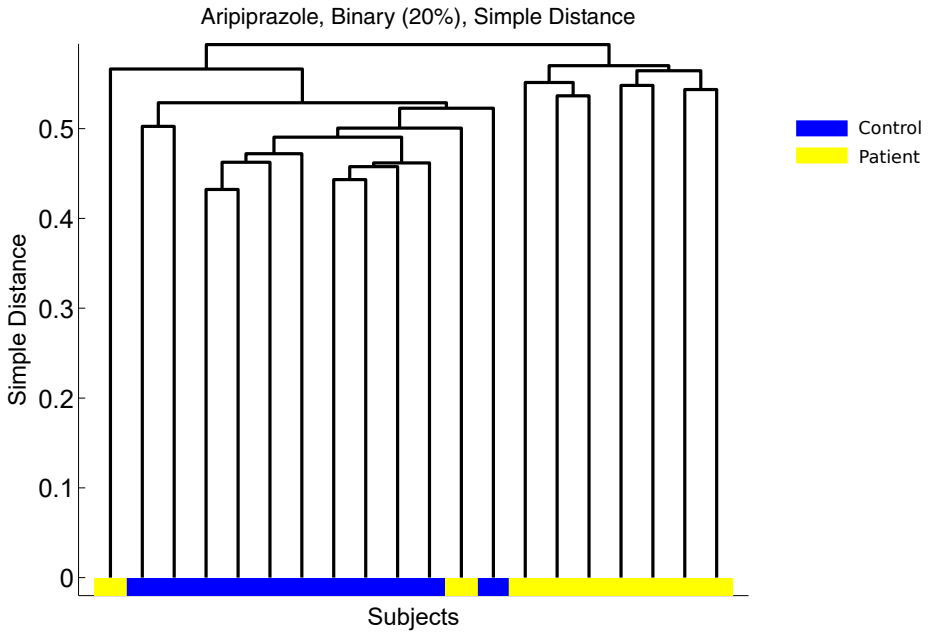


Figure 6.7: Dendrogram for the drug Aripiprazole in which we compare the 11 controls and 9 patients using the distance measure $d_1(\mathbf{A}, \mathbf{B})$. There is some separation between patients and controls.

classify almost every individual. The only exception is Patient 8, who is assigned to the same group as the controls. Although this misclassification seems surprising at first, it agrees with our previous calculations (see Figure 6.4), which also suggest that Patient 8 has different network characteristics than the other patients.

The above result suggests that, under the drug Aripiprazole, we are able to almost completely distinguish patients from controls, based only on information about their community structure. This also suggests that the distance matrix $\mathbf{D}^P$ incorporates more meaningful information than the simplistic distance measure in Equation (6.3), so we use only the former for our subsequent computations.

We show the analogous results comparing controls and patients under placebo in the left panel of Figure 6.9. In this case, we still observe a relatively good separation between patients and controls, in agreement with previous results that functional brain networks encode biomarkers that separate patients diagnosed with schizophrenia from

controls [110, 136]. In this situation as well, Patient 8 appears to be more similar to the controls than to the other patients. Even more interesting, we observe a less-clear separation between the controls and patients than we did under Aripiprazole. We thus conclude that Aripiprazole alters community structure for at least one group and that this alteration makes it easier to distinguish the patient and control groups. However, it is not yet obvious whether Aripiprazole is affecting the architecture of the functional brain networks of patients, controls, or both.

In the right panel of Figure 6.9, we show our results for computations of functional brain networks for individuals under the influence of Sulpiride. The control and patient groups are now less distinct from each other than they were with placebo. This suggests that Sulpiride has a mild but detectable effect of increasing the similarity between community structures of patients and controls. Again, it is not clear whether Sulpiride affects the functional brain networks of patients, controls, or both.

6.4.2 Intra-subject Comparisons

To examine the effects of the drug treatments on network architecture, we make intra-subject comparisons, such as comparing the control group under Aripiprazole to the control group under Sulpiride. We do these comparisons using the procedure that we outlined in Figure 6.1.

6.4.2.1 Aripiprazole versus Placebo

For our intra-subject comparisons (see Figure 6.6), we first compare the effects of Aripiprazole on the functional brain networks of controls to those of patients. To do this, we use all 11 controls under Aripiprazole and the same 11 controls under placebo and do average linkage clustering on the associated 22×22 distance matrix with MRF distances. We also do average linkage clustering using the MRF distance for the 18×18 distance matrix that we obtain by considering the 9 patients under Aripiprazole and the same patients under placebo.

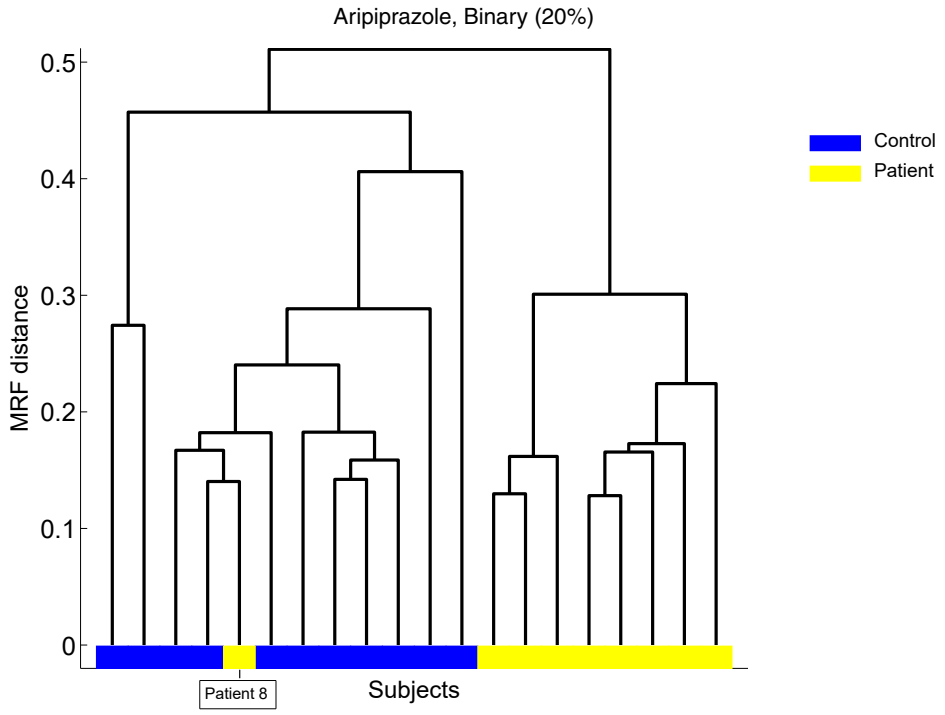


Figure 6.8: Dendrogram for our MRF analysis of functional brain networks for the drug Aripiprazole. We compare the 11 controls and 9 patients using the distance measure $\mathbf{D}^p$. There is a clear separation between patients and controls, although Patient 8 appears with the control group.

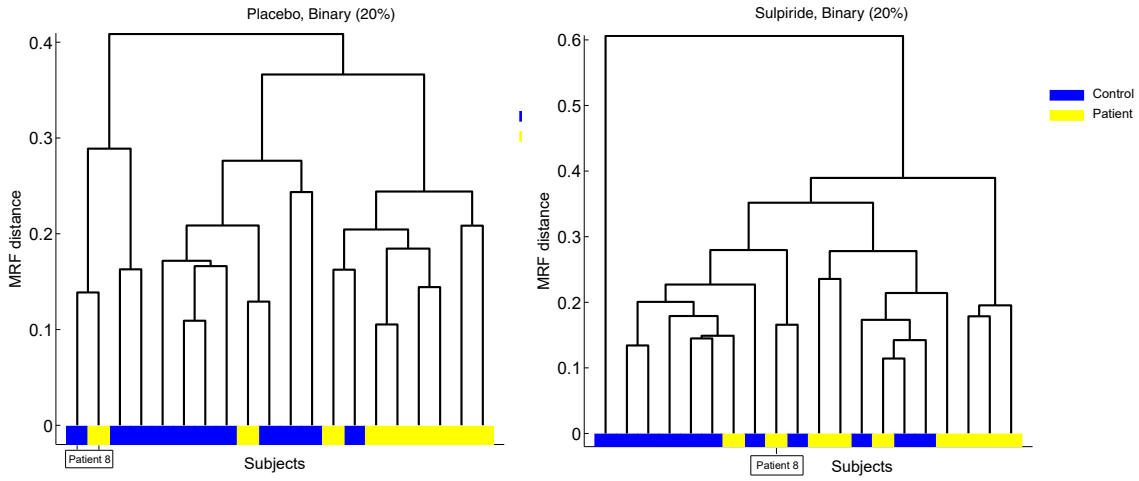


Figure 6.9: Dendrogram for our MRF analysis of functional brain networks for (left) placebo and (right) the drug Sulpiride. In order of most successful to least successful (compare this figure to Figure 6.8), the clustering performs best for Aripiprazole, second-best for placebo, and worst for Sulpiride. (See Section 6.4.3 for a quantitative justification of this observation.)

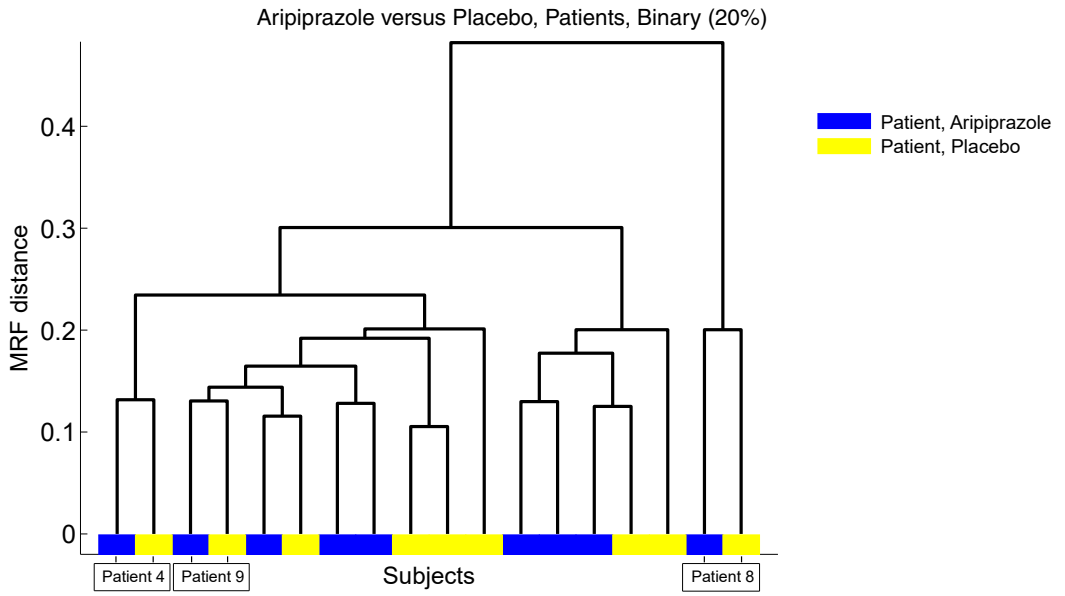


Figure 6.10: Dendrogram for our MRF analysis of functional brain networks for our comparison between Aripiprazole and placebo for the patient group. Each patient thus appears twice on the horizontal axis. There is no clear separation between the two drugs, and the two instances of some patients (e.g., 4, 8, and 9) cluster very close to each other, suggesting that there is very little difference in community structure in the networks under placebo and under Aripiprazole in these patients.

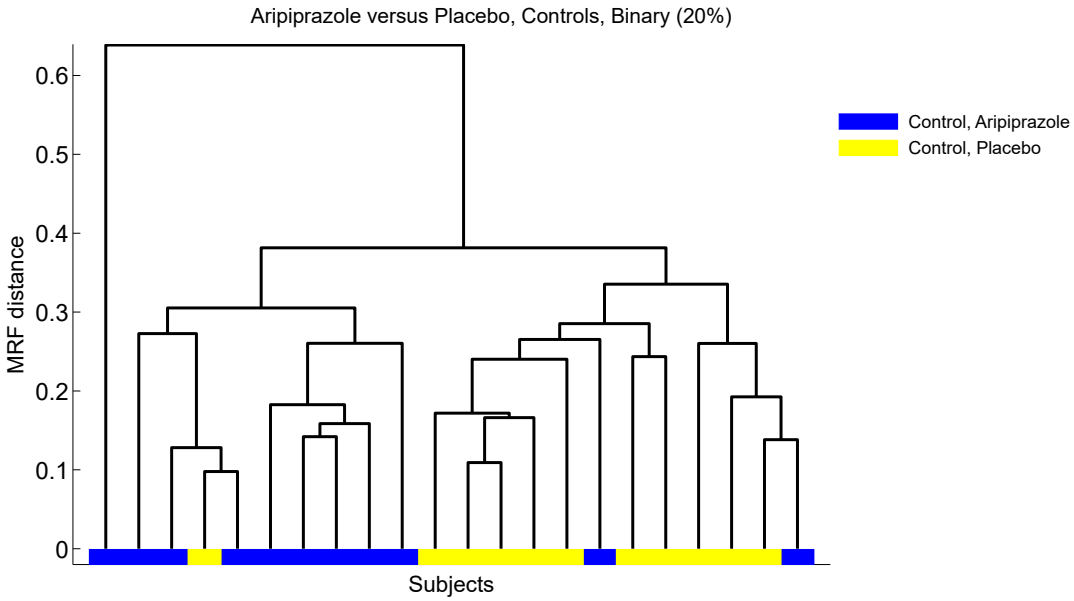


Figure 6.11: Dendrogram for our MRF analysis of functional brain networks for our comparison between Aripiprazole and placebo for the control group. We observe a mostly clear separation between networks under the two drug treatments.

In Figure 6.10, we show the dendrogram for our comparison between Aripiprazole and placebo for patients. At the coarsest level of detail (i.e., a separation for a large MRF distance in the dendrogram), we observe that both the Aripiprazole and placebo network of Patient 8 is grouped away from those of the other patients. This is consistent with our prior results: we saw in Figure 6.4 that Patient 8 has a much larger value of connectivity than the other patients and saw in Figure 6.8 that Patient 8 was grouped with the controls. At the finest level of detail, we also find for both Aripiprazole and placebo that Patients 4 and 9 cluster close to each other. We thus expect, given the inter-subject comparisons in Section 6.4.1, that Aripiprazole does affect community structure in controls. We confirm this hypothesis in Figure 6.11, where we observe that controls under Aripiprazole are clearly separated from controls under placebo.

6.4.2.2 Sulpiride versus Placebo

In Section 6.4.2.1, we observed a very clear separation between controls and patients under the drug Aripiprazole, and we also observed evidence (though the situation is less clear) of separation under placebo. We observed an even lesser separation in Sulpiride. We hypothesized that Sulpiride has a mild but detectable effect of increasing the similarity between community structure in patients and controls, and we therefore hypothesize that Sulpiride affects community structure of either patients or controls (or both), in agreement with [128]. In Figure 6.12, we show a dendrogram of the intra-subject comparison of placebo versus Sulpiride in controls. We do not observe any clear clustering, and we also do not observe clear clustering in the same comparison for patients (see Figure 6.13). We therefore do not find any clear indication of why Sulpiride seems to make controls and patients less distinguishable from each other. Additionally, we do not observe a clear separation under placebo or under Sulpiride either for controls (see Figure 6.12) or for patients (see Figure 6.13).

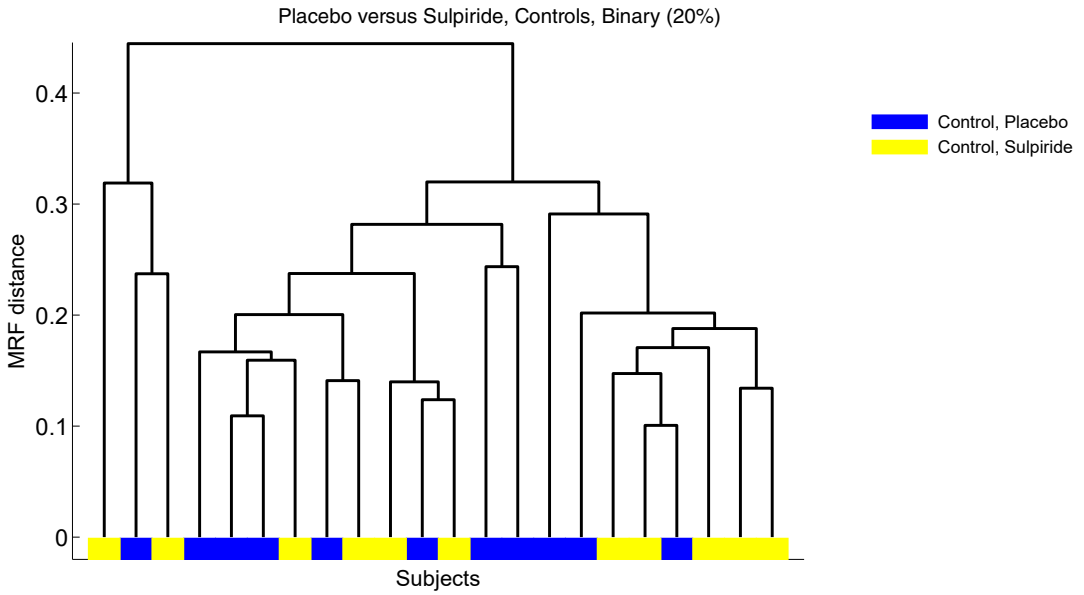


Figure 6.12: Dendrogram for our MRF analysis of functional brain networks for our comparison between Sulpiride and placebo for the control group.

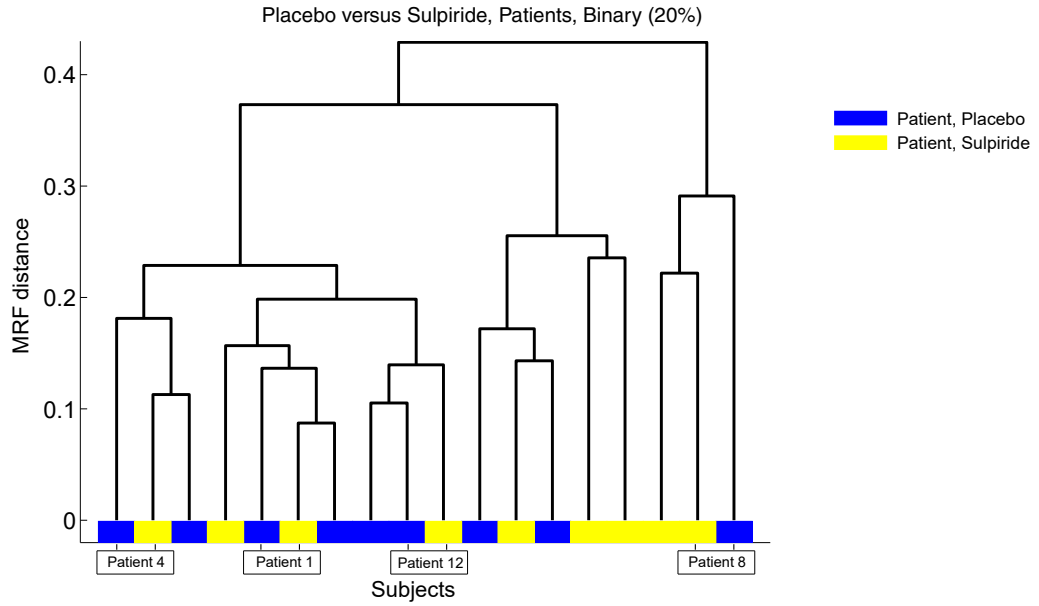


Figure 6.13: Dendrogram for our MRF analysis of functional brain networks for our comparison between Sulpiride and placebo for the patient group. As with our comparison of placebo to Aripiprazole, several identical patients appear close together, and Patient 8 is again distant from the others.

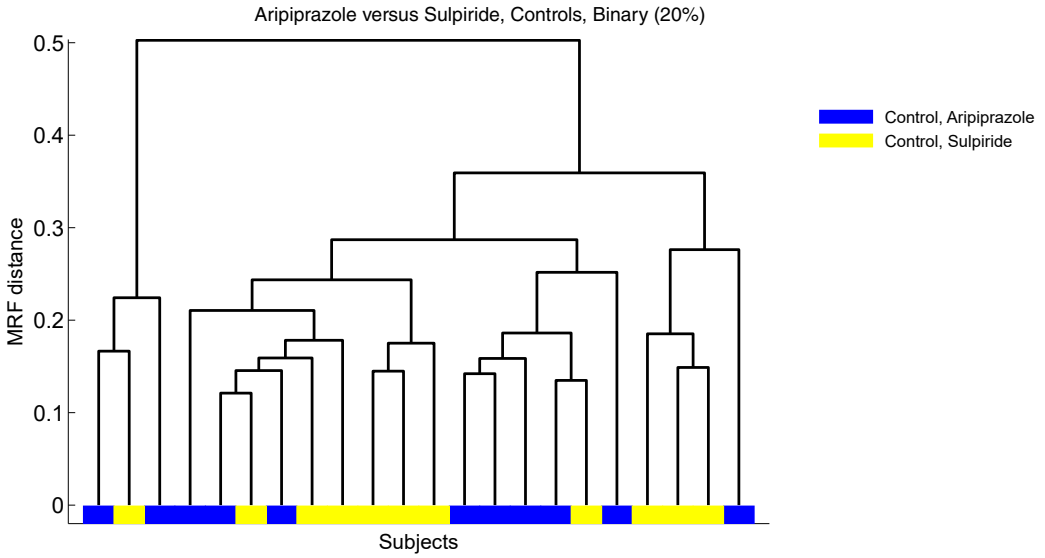


Figure 6.14: Dendrogram for our MRF analysis of functional brain networks for our comparison between Aripiprazole and Sulpiride for the control group.

6.4.2.3 Aripiprazole versus Sulpiride

We can partly distinguish controls under Aripiprazole versus Sulpiride (see Figure 6.14). This is unsurprising, given that we found (see Section 6.4.2.1) that Aripiprazole alters community structure in controls. We do not observe any obvious difference for patients under Aripiprazole versus Sulpiride (see Figure 6.15).

6.4.3 Synthesis of our Results from Hierarchical Clustering and quantitative assessment

Our results from average linkage clustering of collections of functional brain networks using the distance functions yield the following conclusions:

- Aripiprazole affects the community structure of functional brain networks in controls, but not in patients; and it thereby facilitates the distinction between controls and patients under the effect of this drug treatment.

- Sulpiride reduces the distinguishability between patients and controls, although our intra-subject computations were inconclusive in both patients and controls.

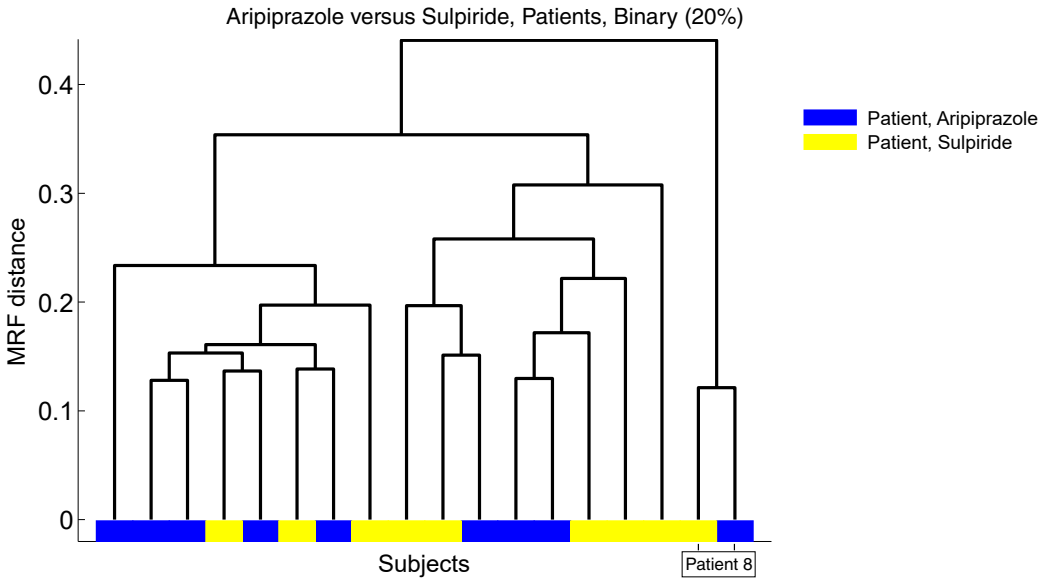


Figure 6.15: Dendrogram for our MRF analysis of functional brain networks for our comparison between Aripiprazole and Sulpiride. In both the Aripiprazole and Sulpiride networks, it is once again easy to distinguish Patient 8 from the other patients.

6.4.3.1 Preliminary quantitative assessment based on purity

In this chapter, we used hierarchical clustering as a simple method to observe how individuals cluster. The main reason for choosing this approach over other possibilities (such as k-means clustering) is that dendrograms provide more information about how individuals group at different distances. For instance, by using dendrograms, we observed in the right panel of Figure 6.10 that Patient 8 is a clear outlier. If we had instead performed k-means clustering with $k = 2$, we would be left with Patient 8 in one community and the other patients in the other community. It would still seem that Patient 8 is an outlier, but the extent to which this is the case would be obscured. Furthermore, hierarchical clustering gives insights at different scales; for example, in the right panel of Figure 6.10, we observed that, at a finer scale, Patient 4 on Aripiprazole is grouped with Patient 4 on placebo. By contrast, at a larger scale, we cannot distinguish communities, which itself is an interesting observation.

The above discussion notwithstanding, it is convenient to attach a number to each par-

tition to quantitatively compare different dendrograms, which is potentially desirable to more precisely evaluate observations such as the extent to which Aripiprazole is better than Sulpiride at separating controls from patients. We leave a detailed analysis for future work, but we perform a preliminary quantitative justification based on the notion of *purity* [150]. Consider a partition of a set of B binary data points (i.e., each data point, which for us is a node, belongs to one of two classes) into k communities. To compute purity, we assign each community to the more-common class in that community, and we measure the accuracy of this assignment by counting the number of correctly-assigned nodes and dividing by B.

To measure the purity of a dendrogram, we use the following simple recipe. For clustering to emerge in a dendrogram, purity should be a nonmonotonic function as a function of the cut level, and it should reach its maximum for a cut at which the number of communities is small (i.e., when the cut is near the top of the dendrogram). In a well-clustered dendrogram, we expect a purity function that peaks at a relatively small number of communities, may then stay roughly constant or decay, and finally increases at the bottom of the dendrogram (as, by definition, purity is trivially maximized when each of the communities has just one component). For well-clustered dendrograms, we take the clustering quality as the earliest peak value of the purity function. However, if purity increases monotonically as a function of the number of communities, we conclude that no good clustering emerges in a dendrogram. For illustration, we sketch the above three typical shapes (which constitute the three simplest nontrivial possibilities) of the purity function in Figure 6.16.

Consider three dendrograms from our inter-subject comparisons: controls versus patients under Aripiprazole (denoted by A; see Figure 6.8), controls versus patients under placebo (denoted by P; see the left panel of Figure 6.9), and controls versus patients under Sulpiride (denoted by S; see the right panel of Figure 6.9). For Sulpiride (S), a dendogram's purity function increases monotonically, so we conclude that there is no good clustering. In other words, one cannot easily distinguish controls from patients

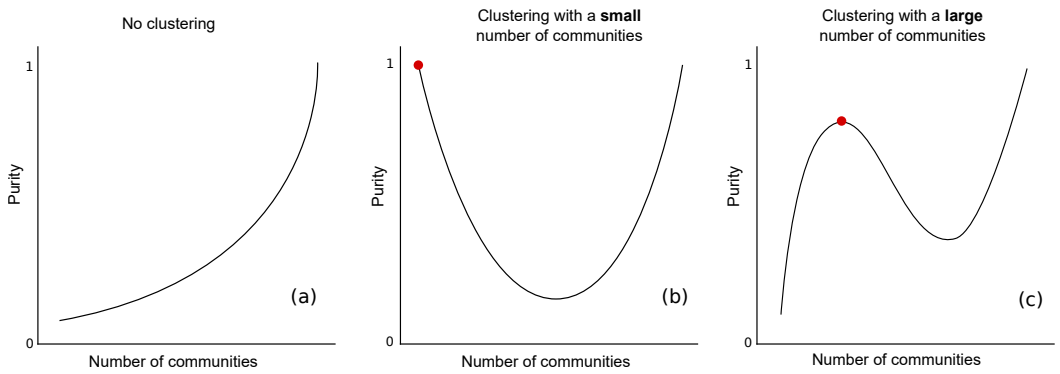


Figure 6.16: Three cartoons that illustrate idealized purity curves as a function of the number of disjoint communities in a dendrogram. In panel (a), purity increases monotonically as a function of the number of communities, so we conclude that no good clustering emerges. In panel (b), purity decreases to a local minimum, and it then increases towards the trivial maximum as the number of communities approaches the total number of nodes. We highlight the clustering quality with a red dot. (We report the purity value at this location as the quality.) In panel (c), purity increases to a local maximum, then decreases to a minimum, and finally approaches the trivial maximum. We again highlight the clustering quality with a red dot.

under Sulpiride, in agreement with our earlier qualitative results discussed. For both Aripiprazole and Sulpiride, however, the dendrogram's purity function is not monotonic, so quantifiable clustering emerges. It peaks at $k = 2$ communities for Aripiprazole (A) and at $k = 3$ communities for placebo (P). The values of purity are

$$\text{purity}(\mathsf{A}) = 19/20 \qquad \text{and} \qquad \text{purity}(\mathsf{P}) = 17/20\,,$$

so observable clustering exists in these two cases, but it is stronger under Aripiprazole than under placebo. This also agrees with our prior qualitative results.

6.4.3.2 An urn-type null model to assess statistical significance

Although our preliminary analysis suggests that there are different levels of clustering in the data, because the data set is small (there are only 20 subjects), one cannot rule out the possibility that putative clustering may be contaminated by statistical artefacts or finite-size effects. A simple way to assess 'how probable' the observed extent of clustering arises from chance is to construct an urn-type null model. We will apply this model to examine the results of the 'controls versus patients under Aripiprazole' dendrogram

(which we labelled as case A in Figure 6.8 of Section 6.4.3). Cutting the dendrogram at the level at which purity peaks yields two communities (each with 10 subjects) and a purity of 19/20. To compute the probability $\mathbb{P}$ that this occurred by chance, we first enumerate all possible partitions. The number Z of different configurations for partitioning the 11 controls and 9 patients into two groups of 10 individuals is

$$Z = 2\sum_{k=0}^{4} \binom{11}{10-k}\binom{9}{k} = 184756\,.$$

There are five possible events:

- Event (E1): There are 10 controls are in one urn and 1 control and 9 patients are in another (which corresponds to case A of controls versus patients under Aripiprazole). The purity is in this case 19/20, and the probability that this event occurs by chance is

  $$\mathbb{P}_{E1} = \frac{2}{Z}\binom{11}{10} = \frac{1}{8398} \approx 10^{-4}\,.$$

  That is, this event would occur randomly with a probability 0.0001, a seemingly very unlikely event.

  For completeness, we also give the estimated probabilities of the other events. We write $\kappa C + \lambda P$ to denote a set with κ controls and λ patients. The probabilities are as follows:

- Event (E2): $9C + 1P$ is in one urn, and $2C + 8P$ is in the other (purity 17/20), so

  $$\mathbb{P}_{E2} = \frac{2}{Z}\binom{11}{9}\binom{9}{1} = \frac{45}{8398} \approx 0.0053\,.$$

- Event (E3): $8C + 2P$ is in one urn, and $3C + 7P$ is in the other (purity 15/20), so

  $$\mathbb{P}_{E3} = \frac{2}{Z}\binom{11}{8}\binom{9}{2} = \frac{270}{4199} \approx 0.064\,.$$

- Event (E4): $7C + 3P$ is in one urn, and $4C + 6P$ is in the other (purity 13/20), so

  $$\mathbb{P}_{E4} = \frac{2}{Z}\binom{11}{7}\binom{9}{3} = \frac{1260}{4199} \approx 0.3\,.$$

- Event (E5): $6C + 4P$ is in one urn, and $5C + 5P$ is in the other (purity 11/20), so

$$\mathbb{P}_{E5} = \frac{2}{Z}\binom{11}{6}\binom{9}{4} = \frac{2646}{4199} \approx 0.63\,.$$

6.5 Discussion

We used network analysis to examine the effects of two antipsychotics, Aripiprazole and Sulpiride, on the architecture of functional brain networks of both controls (who were deemed to be healthy) and patients who have been diagnosed with schizophrenia. Our motivation for our study was twofold: (1) to evaluate whether mesoscale network properties (such as community structure) can distinguish controls from patients diagnosed with schizophrenia and (2) to examine how the results of such calculations different across different types of antipsychotic treatments. Using mesoscopic response functions, we compared community structures of functional brain networks of both patients and controls under the effects of Aripiprazole, Sulpiride, and a placebo.

We will now summarize the results of our computations. However, before doing so, we stress that when interpreting the results of fMRI studies, it is very important to consider the cautionary notes in [151], who noted that computations with fMRI data (even before constructing any networks from such data) rely on a variety of statistical assumptions of questionable validity. These important complications notwithstanding, our computations produced several interesting results. First, we did a reasonable job of distinguishing between controls and patients under placebo. This result suggests that community structure in functional brain networks is a a relevant way to help with diagnoses of schizophrenia. Second, we found that community detection did a much better job (yielding a high-quality clustering) of distinguishing the two groups when Aripiprazole had been administered than for Sulpiride or placebo, suggesting that Aripiprazole has a larger effect on community structure than Sulpiride in at least one of the control group or the patient group. By comparing controls under Aripiprazole and under placebo, we concluded that Aripiprazole appears to improve the distinguishability be-

tween patients and controls through its effects on community structure in the control group, rather than the patient group.

We obtained mixed results for community detection on networks associated with individuals who were treated with the drug Sulpiride. On one hand, we found that patients who were treated with Sulpiride are closer to controls than they are under either Aripiprazole or a placebo (where no clustering seems to emerge, as discussed Section 6.4.3). This is also consistent with [128], and it suggests that Sulpiride has a mild effect on community structure that is appreciably larger than, for instance, the effect of Aripiprazole on community structure in patients (which we observed to be very small). We have not been able to clearly establish the origin of this effect, as our intra-group comparisons suggest that community structure in both controls and patients is mostly unaltered by Sulpiride.

One of the main objectives of an antipsychotic is to manage and reduce symptoms that an individual experiences. For schizophrenia, this involves modifying behaviour and symptoms to cause an individual be more similar, in terms of behaviour and symptoms (or lack thereof), to an average healthy person without the disease. A tempting, but naive, reasoning may suggest that one may therefore expect their associated functional brain networks to also be more similar. This link is poorly understood (though see [128]). Mesoscale network properties, such as communities, are well-known to be important for functional brain networks [101], and network analysis more generally is useful for disentangling structure, function, and their complex interrelations in the brain. However, the link between drug effectiveness and the effect on functional brain networks is not clear and both merits and requires further investigation. It is noteworthy that our observations that Aripiprazole primarily affects community structure in controls, rather than patients, is consistent with the results of [128], who reported that Aripiprazole has a radical effect on the organization of the brain networks of controls but decreases the performance of the controls at cognitive tasks. Our work also leaves additional open questions for future work. For instance, an interesting technical point that is worth exploring in more detail is to examine clustering methods other than hierarchical clustering. The sample size

(20 subjects) in the experimental data that we studied is small, and conducting new, large-scale experiments is highly desirable to test the validity of our results (although evidence in Section 6.4.3 and Section 6.4.3.2 suggests that our results are statistically significant).

From a network-science viewpoint, we highlight that we used community structure and mesoscopic response functions for a classification task in time-independent, monolayer functional brain networks. Extending these results and analysis to time-dependent and multilayer settings [152, 153] is another interesting open problem.

6.6 Appendix: Network Component Sizes

As we discussed in Section 6.3.2, half of our networks (30 out of 60) consist of two or more components after thresholding. However, even in these cases, the largest connected component of each network includes almost the entire network. In Table 6 in Appendix 6.6, at the bottom of this chapter, we show the number of components and component sizes for each of the 60 networks.

Table 6: Number of components and component sizes of each of the 60 networks. We denote treatment under Aripiprazole by "A", treatment under Sulpiride by "S", and treatment under placebo by "P".

Subject	*Number of Components*	*Component Sizes*
Control 1 (A)	6	{293,1,1,1,1,1}
Control 3 (A)	4	{295,1,1,1}
Control 4 (A)	5	{293,2,1,1,1}
Control 5 (A)	20	{268,6,5,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1}
Control 6 (A)	2	{297,1}
Control 7 (A)	3	{296,1,1}
Control 9 (A)	4	{295,1,1,1}
Control 11 (A)	2	{297,1}
Control 12 (A)	6	{292,2,1,1,1,1}

Control 13 (A)	6	{292,2,1,1,1,1}
Control 15 (A)	2	{297,1}
Control 1 (P)	1	298
Control 3 (P)	2	{297,1}
Control 4 (P)	1	298
Control 5 (P)	2	{297,1}
Control 6 (P)	5	{289,3,3,2,1}
Control 7 (P)	4	{295,1,1,1}
Control 9 (P)	2	{297,1}
Control 11 (P)	1	298
Control 12 (P)	4	{294,2,1,1}
Control 13 (P)	4	{295,1,1,1}
Control 15 (P)	1	298
Control 1 (S)	1	298
Control 3 (S)	2	{297,1}
Control 4 (S)	1	298
Control 5 (S)	1	298
Control 6 (S)	3	{296,1,1}
Control 7 (S)	1	298
Control 9 (S)	1	298
Control 11 (S)	2	{297,1}
Control 12 (S)	1	298
Control 13 (S)	3	{296,1,1}
Control 15 (S)	1	298
Patient 1 (A)	2	{297,1}
Patient 2 (A)	1	298
Patient 4 (A)	1	298
Patient 6 (A)	1	298

Patient 7 (A)	2	{297,1}
Patient 8 (A)	7	{291,2,1,1,1,1,1}
Patient 9 (A)	1	298
Patient 10 (A)	1	298
Patient 12 (A)	1	298
Patient 1 (P)	1	298
Patient 2 (P)	1	298
Patient 4 (P)	1	298
Patient 6 (P)	1	298
Patient 7 (P)	2	{297,1}
Patient 8 (P)	7	{292,1,1,1,1,1,1}
Patient 9 (P)	1	298
Patient 10 (P)	3	{296,1,1}
Patient 12 (P)	1	298
Patient 1 (S)	1	298
Patient 2 (S)	2	{297,1}
Patient 4 (S)	1	298
Patient 6 (S)	1	298
Patient 7 (S)	1	298
Patient 8 (S)	5	{294,1,1,1,1}
Patient 9 (S)	1	298
Patient 10 (S)	1	298
Patient 12 (S)	1	298

Chapter 7

Conclusion

In this thesis, we have used network science and graph theory to deal with time series data in a range of scenarios. In retrospect, all our results could be classified in two categories: theoretical advancements, and new applications and their implications. The theoretical advancements deal with new mathematical and methodological results, which have value in their own right, whereas the latter deal with new answers to specific questions which arise when we apply these methods to different areas outside mathematics.

Chapters 2 to 5 all involve the use of visibility graphs: either applying existing methods to new data (such as Chapter 3), proposing new methods (such as Chapter 4), and deriving some mathematical results on the structure (such as Chapter 5. Chapter 6 uses a different approach; estimating correlation networks from fMRI data and using the community structure of these networks to make an analysis of a brain condition (schizophrenia) and two drugs. In what follows we discuss the specific results obtained in each of the Chapters, to which category they belong to, and a list of open problems.

In Chapter 2, we used the definitions of natural and horizontal visibility graphs to and a method using the Kullback-Leibler divergence (between the in and out degree distributions of the visibility graph) to quantify degrees of time irreversibility in non-stationary processes. We proved that unbiased additive random walks, while non-stationary, are

both VG/HVG-stationary and VG/HVG-time reversible. On the other hand, we found that biased memoryless additive random walks are HVG-irreversible with finite irreversibility measures that quantify the degree of time asymmetry, and that they are still VG-reversible. Our numerics suggested that HVG can capture irreversibility for both finite and infinite series size, whereas VG is only able to do so for finite series. We found that multiplicative random walks, the processes are HVG-reversible if the process is akin to an unbiased additive process in logarithmic space, and time irreversible if the process reduces to a biased additive process in logarithmic space.

In Chapter 3, we applied this new theory and the associated method (using KLD to capture irreversibility in non-stationary processes) to a large financial data set. We looked at financial market stability and whether if it is in 'thermodynamic equilibrium'. We asked if 'thermodynamic equilibrium' (i) can can be investigated, (ii) is related to time series irreversibility, (iii) can for the first time be explored quantitatively using visibility graph formalism. We found that the stock prices of the companies in our dataset are indeed time irreversible. This is yet more evidence that exposes the inefficiency of financial systems, supporting the violation of the classical efficient market hypothesis. We note that different companies had distinct time evolving irreversibility patterns, and some displayed periods of quasi-reversibility, which implies that some companies are more irreversible than others, and the degree of reversibility of each company varies over time. We also ranked the companies based on their irreversibility (our "Score" metric from Section 3.3.2, and according to the analogy between reversibility and entropy production, the larger this score is, the more 'away from equilibrium' the signal generated by the company is, thus producing larger amounts of entropy. We thus demonstrated a method for quantifying irreversibility in non-stationary time series by applying it to real world data, revealing complementary (to existing methods) and valuable information on the evolution and the structure of stock prices. Further research is needed to explore the relation between irreversibility, predictability, or more generally, the evolution and dynamics of stock prices.

In Chapter 4, we laid down the framework to quantify peak/pit asymmetries in complex systems. We noted that some stochastic processes are statistically invariant under the inversion $x(t) \to -x(t)$. It is interesting to understand which of these processes have natural and horizontal visibility graphs which themselves are statistically invariant under such an inversion. In Chapters 2 and 3 our analysis was done to the original series $x(t)$, which in general relates to the structure of the maxima of a signal, at the expense of the minima. However, in many applications, the dynamics which originate the maxima and minima are different (in neuroscience for example). Our framework provides a first look into this problem from the perspective of visibility graphs. We use a variety of real world data to validate our new methods, and while the results certified that the method works as it should, no new or unexpected insights have been obtained for these datasets. There is huge potential to apply our method to noisy real world data to capture information in both the maxima and the minima of the relating time series.

In Chapter 5, we derived the first theoretic results on spectral properties of horizontal visibility graphs relating to the Feigenbaum scenario. We were able to enumerate the visibility graphs below the logistic map's accumulation point in terms of a bi-parametric family of finite Feigenbaum graphs F_n^k, and we explored their spectral properties.We found noteworthy patterns, and our numerical results were complemented with analytical developments as well as exact results (such as the diameter of F_n and the determinant of F_n^k). Other aspects that were investigated include the full spectrum, the number of distinct eigenvalues, and the number of spanning trees of the whole family of F_n^k. A similar analysis was then conducted in the region of the map's parameter where trajectories are chaotic. We found that the maximal eigenvalue, while being a good discriminator between chaos and noise, is not able to quantify chaoticity, thereby challenging the previously accepted assumption that Graph Index Complexity is a general indicator of the time series complexity. In this sense, we found that the eigenvalue distribution carried more information about time series chaoticity, in particular its Lyapunov exponent. Some open problems include exploring the fractal-like shape the full spectrum of F_n^k ap-

proaches as n increases, and aiming to improve the bounds on $\lambda_{\max}$ for F_n^k. In the case of $\mu = \mu_\infty$, the Feigenbaum graph F_∞^∞ is no longer locally finite. To be able to relate its spectral properties to its dynamical counterpart also constitutes an intriguing problem. For the region in $\mu > \mu_\infty$, where the map is chaotic, the connection between the spectral properties of the associated graphs and the dynamical properties of this unimodal map has only been explored numerically, and therefore stating these connections (or the lack of) remain as an open problem.

In Chapter 6, we use an established method to construct correlation networks from fMRI data. We then used a recent method to analyse the community structure of these networks by seeing how the communities change when varying the resolution parameter. We did not discuss any new theory (aside from using some statistics to choose the scale and thresholding parameter instead of arbitrarily selecting it, and proposing a method using purity to examine the success of clustering methods). We first showed that the distance measure based on community structure is better at distinguishing the two groups of patients and controls compared to a basic distance measure. We then showed that the clustering between patients and controls is larger in some drugs than others, and also the clustering between one group and themselves is larger in some drugs than others. This suggests that with regards to human brain fMRI data, community structure is an informative and relevant mesoscale structure which in theory could be useful for diagnosing a particular disease. It could also be useful in measuring the effect certain drugs have on the brain. It is noteworthy that in our observations, Aripiprazole primarily affects community structure in controls, rather than patients (and is consistent with the results of [128]). This poses an open question, which is "What is the effect drugs have on the brains of subjects for whom the drug is not designed for?". More would should be done to probe this effect. Furthermore, an interesting technical point that is worth exploring in more detail is to examine methods (such as the one we outlined) to quantitatively assess qualitative methods such hierarchical clustering, and to examine clustering methods other than this one.

In summary, we have shown how network theory can be used to study data. We have done this with a variety of methods, from several angles, with data from several fields. We have also displayed the usefulness of interdisciplinary research in studying this data, as several pools of knowledge can be combined to answer a wider variety of questions.

Bibliography

[1] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths. "Complex network approaches to nonlinear time series analysis." *Physics Reports*, 787:1–97, 2019.

[2] M. De Berg, O. Cheong, M. Van Kreveld, and M. Overmars. *Computational geometry: Algorithms and applications*. Springer Berlin Heidelberg, 2008. ISBN 9783540779735.

[3] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño. "From time series to complex networks: The visibility graph." *Proceedings of the National Academy of Sciences of the United States of America*, 105(13):4972–4975, 2008.

[4] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque. "Horizontal visibility graphs: Exact results for random time series." *Physical Review E* -, 80(4):046103, 2009.

[5] L. Lacasa and R. Flanagan. "Time reversibility from visibility graphs of nonstationary processes." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 92(2):022817, 2015.

[6] R. Flanagan and L. Lacasa. "Irreversibility of financial time series: A graph-theoretical approach." *Physics Letters A*, 380(20):1689–1697, 2016.

[7] U. Hasson, J. Iacovacci, B. Davis, R. Flanagan, E. Tagliazucchi, H. Laufs, and L. Lacasa. "A combinatorial framework to quantify peak/pit asymmetries in complex dynamics." *Scientific Reports*, 8(1), 2018.

[8] R. Flanagan, L. Lacasa, and V. Nicosia. "On the spectral properties of Feigenbaum graphs." *Journal of Physics A: Mathematical and Theoretical*, 53(2):025702, 2020.

[9] R. Flanagan, L. Lacasa, E. K. Towlson, S. H. Lee, and M. A. Porter. "Effect of antipsychotics on community structure in functional brain networks." *Journal of Complex Networks*, 7(6):932–960, 2019.

[10] A. J. Lawrance. "Directionality and Reversibility in Time Series." *International Statistical Review / Revue Internationale de Statistique, Vol. 59*, 596949135(1):67–79, 1991.

[11] C. S. Daw, C. E. Finney, and M. B. Kennel. "Symbolic approach for measuring temporal "irreversibility"." *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 62(2 B):1912–1921, 2000.

[12] É. Roldán and J. M. R. Parrondo. "Estimating Dissipation from Single Stationary Trajectories." *Physical Review Letters*, 105(15):150607, 2010.

[13] É. Roldán and J. M. Parrondo. "Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems." *Physical Review E*, 85(3):031129, 2012.

[14] M. Costa, A. L. Goldberger, and C. K. Peng. "Broken asymmetry of the human heartbeat: Loss of time irreversibility in aging and disease." *Physical Review Letters*, 95(19):198102, 2005.

[15] C. Cammarota and E. Rogora. "Time reversal, symbolic series and irreversibility of human heartbeat." *Chaos, Solitons and Fractals*, 32(5):1649–1654, 2007.

[16] M. B. Kennel. "Testing time symmetry in time series using data compression dictionaries." *Physical Review E*, 69(5):9, 2004.

[17] G. Weiss. "Time-Reversibility of Linear Stochastic Processes." *Journal of Applied Probability*, 12(4):831–836, 1975.

[18] L. Lacasa, A. Nuñez, E. Roldán, J. M. Parrondo, and B. Luque. "Time series irreversibility: A visibility graph approach." *European Physical Journal B*, 85(6):217, 2012.

[19] J. F. Donges, R. V. Donner, and J. Kurths. "Testing time series irreversibility using complex network methods." *Epl*, 102(1):10004, 2013.

[20] V. Suyal, A. Prasad, and H. P. Singh. "Visibility-Graph Analysis of the Solar Wind Velocity." *Solar Physics*, 289(1):379–389, 2014.

[21] F. Xie, Z. Fu, L. Piao, and J. Mao. "Time irreversibility of mean temperature anomaly variations over China." *Theoretical and Applied Climatology*, 123(1-2):161–170, 2016.

[22] C. F. Schleussner, D. V. Divine, J. F. Donges, A. Miettinen, and R. V. Donner. "Indications for a North Atlantic ocean circulation regime shift at the onset of the Little Ice Age." *Climate Dynamics*, 45(11-12):3623–3633, 2015.

[23] Y. Zou, R. V. Donner, N. Marwan, M. Small, and J. Kurths. "Long-term changes in the north-south asymmetry of solar activity: A nonlinear dynamics characterization using visibility graphs." *Nonlinear Processes in Geophysics*, 21(6):1113–1126, 2014.

[24] Sidney Redner. *A Guide to First-Passage Processes*. 2001.

[25] L. Lacasa, B. Luque, J. Luque, and J. C. Nũo. "The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion." *Epl*, 86(3):30001, 2009.

[26] B. Luque, L. Lacasa, F. J. Ballesteros, and A. Robledo. "Feigenbaum graphs: A complex network perspective of chaos." *PLoS ONE*, 6(9):e22411, 2011.

[27] Y. Long. "Visibility graph network analysis of gold price time series." *Physica A: Statistical Mechanics and its Applications*, 392(16):3374–3384, 2013.

[28] C. Liu, W.-X. X. Zhou, and W.-K. K. Yuan. "Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence." *Physica A: Statistical Mechanics and its Applications*, 389(13):2675–2681, 2010.

[29] M. Murugesan and R. I. Sujith. "Combustion noise is scale-free: Transition from scale-free to order at the onset of thermoacoustic instability." *Journal of Fluid Mechanics*, 772:225–245, 2015.

[30] M. Ahmadlou, H. Adeli, and A. Adeli. "New diagnostic EEG markers of the Alzheimer's disease using visibility graph." *Journal of Neural Transmission*, 117(9):1099–1109, 2010.

[31] M.-C. Qian, Z.-Q. Jiang, and W.-X. Zhou. "Universal and nonuniversal allometric scaling behaviors in the visibility graphs of world stock market indices." *Journal of Physics A: Mathematical and Theoretical*, 43(33):335002, 2010.

[32] N. Wang, D. Li, and Q. Wang. "Visibility graph analysis on quarterly macroeconomic series of China based on complex network theory." *Physica A: Statistical Mechanics and its Applications*, 391(24):6543–6555, 2012.

[33] L. Lacasa and R. Toral. "Description of stochastic and chaotic series using visibility graphs." *Physical Review E*, 82(3):036120, 2010.

[34] Á. M. Núñez, B. Luque, L. Lacasa, J. P. Gómez, and A. Robledo. "Horizontal visibility graphs generated by type-I intermittency." *Physical Review E*, 87(5):052801, 2013.

[35] B. Luque, F. J. Ballesteros, A. M. Nunez, A. Robledo, A. M. Núñez, and A. Robledo. "Quasiperiodic Graphs: Structural Design, Scaling and Entropic Properties." *Journal of Nonlinear Science*, 23(2):335–342, 2013.

[36] L. Lacasa. "On the degree distribution of horizontal visibility graphs associated with Markov processes and dynamical systems: Diagrammatic and variational approaches." *Nonlinearity*, 27(9):2063–2093, 2014.

[37] D. Boyer and C. Solis-Salas. "Random walks with preferential relocations to places visited in the past and their application to biology." *Physical Review Letters*, 112(24):240601, 2014.

[38] S. Redner. "Random multiplicative processes: An elementary tutorial." *American Journal of Physics*, 58(3):267–273, 1990.

[39] R. S. Tsay. *Analysis of financial time series*. Wiley, 2005. ISBN 0471746185.

[40] H. E. Stanley, L. A. Amaral, D. Canning, P. Gopikrishnan, Y. Lee, and Y. Liu. "Econophysics: Can physicists contribute to the science of economics?" *Physica A: Statistical Mechanics and its Applications*, 269(1):156–169, 1999.

[41] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Amaral, and H. Eugene Stanley. "Econophysics: Financial time series from a statistical physics point of view." *Physica A: Statistical Mechanics and its Applications*, 279(1):443–456, 2000.

[42] R. N. Mantegna and H. E. Stanley. "Introduction to Econophysics." *Cambridge University Press*, 1999.

[43] T. Di Matteo, T. Aste, and M. M. Dacorogna. "Scaling behaviors in differently developed markets." *Physica A: Statistical Mechanics and its Applications*, 324(1-2):183–188, 2003.

[44] T. Di Matteo, T. Aste, and M. M. Dacorogna. "Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development." *Journal of Banking and Finance*, 29(4):827–851, 2005.

[45] D. O. Cajueiro and B. M. Tabak. "The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient." *Physica A: Statistical Mechanics and its Applications*, 336(3-4):521–537, 2004.

[46] D. O. Cajueiro and B. M. Tabak. "Ranking efficiency for emerging markets." *Chaos, Solitons & Fractals*, 22(2):349–352, 2004.

[47] C. Eom, S. Choi, G. Oh, and W. S. Jung. "Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets." *Physica A: Statistical Mechanics and its Applications*, 387(18):4630–4636, 2008.

[48] L. Zunino, B. M. Tabak, A. Figliola, D. G. Pérez, M. Garavaglia, and O. A. Rosso. "A multifractal approach for stock market inefficiency." *Physica A: Statistical Mechanics and its Applications*, 387(26):6558–6566, 2008.

[49] L. Zunino, M. Zanin, B. M. Tabak, D. G. Pérez, and O. A. Rosso. "Forbidden patterns, permutation entropy and stock market inefficiency." *Physica A: Statistical Mechanics and its Applications*, 388(14):2854–2864, 2009.

[50] C. Eom, G. Oh, and W.-S. Jung. "Relationship between efficiency and predictability in stock price change." *Physica A: Statistical Mechanics and its Applications*, 387(22):5511–5517, 2008.

[51] J. Ramsey and P. Rothman. "Time irreversibility and business cycle asymmetry." *Journal of Money, Credit and Banking*, 28(1):1–21, 1996.

[52] Y.-T. Chen, R. Y. Chou, and C.-M. Kuan. "Testing time reversibility without moment restrictions." *Journal of Econometrics*, 95(1):199–218, 2000.

[53] W. M. Fong. "Time reversibility tests of volume-volatility dynamics for stock returns." *Economics Letters*, 81(1):39–45, 2003.

[54] G. Zumbach. "Time reversal invariance in finance." *Quantitative Finance*, 9(5):505–515, 2009.

[55] J. Xia, P. Shang, J. Wang, and W. Shi. "Classifying of financial time series based on multiscale entropy and multiscale time irreversibility." *Physica A: Statistical Mechanics and its Applications*, 400:151–158, 2014.

[56] E. Martina, E. Rodriguez, R. Escarela-Perez, and J. Alvarez-Ramirez. "Multiscale entropy analysis of crude oil price dynamics." *Energy Economics*, 33(5):936–947, 2011.

[57] N. M. Nasrabadi. "Pattern Recognition and Machine Learning." *Journal of Electronic Imaging*, 16(4):049901, 2007.

[58] L. Lacasa, V. Nicosia, and V. Latora. "Network structure of multivariate time series." *Scientific Reports*, 5(1):15508, 2015.

[59] H. Kantz and T. Schreiber. *Nonlinear time series analysis*. Cambridge University Press, 2003. ISBN 0521529026.

[60] D. Cordes, V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand. "Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data." *American Journal of Neuroradiology*, 22(7):1326–1333, 2001.

[61] N. Petridou, C. C. Gaudes, I. L. Dryden, S. T. Francis, and P. A. Gowland. "Periods of rest in fMRI contain individual spontaneous events which are related to slowly fluctuating spontaneous activity." *Human Brain Mapping*, 34(6):1319–1329, 2013.

[62] E. Tagliazucchi, M. Siniatchkin, H. Laufs, and D. R. Chialvo. "The voxel-wise functional connectome can be efficiently derived from co-activations in a sparse spatio-temporal point-process." *Frontiers in Neuroscience*, 10(AUG):381, 2016.

[63] X. Liu and J. H. Duyn. "Time-varying functional network information extracted from brief instances of spontaneous brain activity." *Proceedings of the National Academy of Sciences*, 110(11):4392–4397, 2013.

[64] A. D. Patel and E. Balaban. "Temporal patterns of human cortical activity reflect tone sequence structure." *Nature*, 404(6773):80–84, 2000.

[65] A. Mazaheri and O. Jensen. "Rhythmic pulsing: Linking ongoing brain activity with evoked responses." *Frontiers in Human Neuroscience*, 4:177, 2010.

[66] B. Davis, J. Jovicich, V. Iacovella, and U. Hasson. "Functional and Developmen-

tal Significance of Amplitude Variance Asymmetry in the BOLD Resting-State Signal." *Cerebral Cortex*, 24(5):1332–1350, 2014.

[67] B. Davis, E. Tagliazucchi, J. Jovicich, H. Laufs, and U. Hasson. "Progression to deep sleep is characterized by changes to BOLD dynamics in sensory cortices." *NeuroImage*, 130:293–305, 2016.

[68] M. G. Ravetti, L. C. Carpi, B. A. Gonçalves, A. C. Frery, and O. A. Rosso. "Distinguishing noise from chaos: Objective versus subjective criteria using horizontal visibility graph." *PLoS ONE*, 9(9):e108004, 2014.

[69] J. Iacovacci and L. Lacasa. "Sequential visibility-graph motifs." *Physical Review E*, 93(4):042309, 2016.

[70] T. A. Amor, R. Russo, I. Diez, P. Bharath, M. Zirovich, S. Stramaglia, J. M. Cortes, L. De Arcangelis, and D. R. Chialvo. "Extreme brain events: Higher-order statistics of brain resting activity and its relation with structural connectivity." *Epl*, 111(6):68007, 2015.

[71] Y. Zou, M. Small, Z. Liu, and J. Kurths. "Complex network approach to characterize the statistical features of the sunspot series." *New Journal of Physics*, 16(1):013051, 2014.

[72] B. Luque and L. Lacasa. "Canonical horizontal visibility graphs are uniquely determined by their degree sequence." *European Physical Journal: Special Topics*, 226(3):383–389, 2017.

[73] R. M. Birn, J. B. Diamond, M. A. Smith, and P. A. Bandettini. "Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI." *NeuroImage*, 31(4):1536–1548, 2006.

[74] Y. Zhang, M. Brady, and S. Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm." *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.

[75] D. N. Greve and B. Fischl. "Accurate and robust brain image alignment using boundary-based registration." *NeuroImage*, 48(1):63–72, 2009.

[76] S. D. Forman, J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, and D. C. Noll. "Improved Assessment of Significant Activation in Functional Magnetic Resonance Imaging (fMRI): Use of a Cluster-Size Threshold." *Magnetic Resonance in Medicine*, 33(5):636–647, 1995.

[77] M. Kanamitsu, W. Ebisuzaki, J. Woollen, S. K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter. "NCEP-DOE AMIP-II reanalysis (R-2)." *Bulletin of the American Meteorological Society*, 83(11):1631–1643+1559, 2002.

[78] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[79] J. Kim and T. Wilhelm. "What is a complex graph?" *Physica A: Statistical Mechanics and its Applications*, 387(11):2637–2652, 2008.

[80] X. Tang, L. Xia, Y. Liao, W. Liu, Y. Peng, T. Gao, and Y. Zeng. "New approach to epileptic diagnosis using visibility graph of high-frequency signal." *Clinical EEG and Neuroscience*, 44(2):150–156, 2013.

[81] M. Mozaffarilegha and H. Adeli. "Visibility graph analysis of speech evoked auditory brainstem response in persistent developmental stuttering." *Neuroscience Letters*, 696:28–32, 2019.

[82] M. Nasrolahzadeh, Z. Mohammadpoory, and J. Haddadnia. "Analysis of heart rate signals during meditation using visibility graph complexity." *Cognitive Neurodynamics*, 2019.

[83] V. Fioriti, A. Tofani, and A. Di Pietro. "Discriminating chaotic time series with visibility graph eigenvalues." Technical Report 3, 2012.

[84] B. Luque, L. Lacasa, F. J. Ballesteros, and A. Robledo. "Analytical properties of

horizontal visibility graphs in the Feigenbaum scenario." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(1):13109, 2012.

[85] R. J. Wilson. *Introduction to Graph Theory*. Longman, 1979.

[86] Y. Shang. "Geometric assortative growth model for small-world networks." *The Scientific World Journal*, 2014.

[87] B. Mohar and W. Woess. "A survey on spectra of infinite graphs." *Bulletin of the London Mathematical Society*, 21(3):209–234, 1989.

[88] S. G. Hwang. "Cauchy's Interlace Theorem for Eigenvalues of Hermitian Matrices." *American Mathematical Monthly*, 111(2):157–159, 2004.

[89] K. C. Das and P. Kumar. "Some new bounds on the spectral radius of graphs." *Discrete Mathematics*, 281(1):149–161, 2004.

[90] P. V. Mieghem. *Graph spectra for complex networks*, volume 9780521194. Cambridge University Press, 2010. ISBN 9780511921681.

[91] N. Biggs. *Algebraic graph theory*. Cambridge university press, 1993.

[92] F. Harary. "The Determinant of the Adjacency Matrix of a Graph." *SIAM Review*, 4(3):202–210, 1962.

[93] A. Aharony and A. B. Harris. "Absence of self-averaging and universal fluctuations in random systems near critical points." *Physical Review Letters*, 77(18):3700–3703, 1996.

[94] S. H. Strogatz. *Nonlinear dynamics and chaos*. Biology, Chemistry, and Engineering. 1st ed Perseus Books, 1994.

[95] L. Lacasa and W. Just. "Visibility graphs and symbolic dynamics." *Physica D: Nonlinear Phenomena*, 374-375:35–44, 2018.

[96] S. L. Braunstein, S. Ghosh, and S. Severini. "The Laplacian of a graph as a density Matrix: A basic combinatorial approach to separability of mixed states." *Annals of Combinatorics*, 10(3):291–317, 2006.

[97] L. A. Bunimovich and B. Z. Webb. "Isospectral graph reductions and improved estimates of matrices' spectra." *Linear Algebra and Its Applications*, 437(7):1429–1457, 2012.

[98] A. Francis, D. Smith, D. Sorensen, and B. Webb. "Extensions and applications of equitable decompositions for graphs with symmetries." *Linear Algebra and Its Applications*, 532:432–462, 2017.

[99] B. Luque, L. Lacasa, and A. Robledo. "Feigenbaum graphs at the onset of chaos." *Physics Letters A*, 376(47):3625–3629, 2012.

[100] D. S. Bassett and O. Sporns. "Network neuroscience." *Nature Neuroscience*, 20(3):353–364, 2017.

[101] R. F. Betzel and D. S. Bassett. "Multi-scale brain networks." *NeuroImage*, 160:73–83, 2017.

[102] O. Sporns, G. Tononi, and R. Kötter. "The human connectome: A structural description of the human brain." *PLoS Computational Biology*, 1(4):0245–0251, 2005.

[103] O. Sporns. "Structure and function of complex brain networks." *Dialogues in Clinical Neuroscience*, 15(3):247–262, 2013.

[104] O. Sporns. "Cerebral cartography and connectomics." *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1668):20140173, 2015.

[105] O. Sporns and R. F. Betzel. "Modular brain networks." *Annual Review of Psychology*, 67(1):613–640, 2015.

[106] D. Papo, J. M. Buldú, S. Boccaletti, and E. T. Bullmore. "Complex network theory and the brain." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1653):20130520, 2014.

[107] S. E. Petersen and O. Sporns. "Brain Networks and Cognitive Architectures." *Neuron*, 88(1):207–219, 2015.

[108] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius. "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease." *PLoS Computational Biology*, 4(6):e1000100, 2008.

[109] G. S. Dichter. "Functional magnetic resonance imaging of autism spectrum disorders." *Dialogues in clinical neuroscience*, 14(3):319–351, 2012.

[110] M.-E. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, and E. T. Bullmore. "Functional connectivity and brain networks in schizophrenia." *The Journal of Neuroscience*, 30(28):9477–9487, 2010.

[111] M. P. Van Den Heuvel, O. Sporns, G. Collin, T. Scheewe, R. C. W. Mandl, W. Cahn, J. Goñi, H. E. Hulshoff Pol, R. S. Kahn, J. Goni, H. E. Pol, and R. S. Kahn. "Abnormal rich club organization and functional brain dynamics in schizophrenia." *JAMA Psychiatry*, 70(8):783–792, 2013.

[112] J. A. Hadley, N. V. Kraguljac, D. M. White, L. Ver Hoef, J. Tabora, and A. C. Lahti. "Change in brain network topology as a function of treatment response in schizophrenia: A longitudinal resting-state fMRI study using graph theory." *npj Schizophrenia*, 2:16014, 2016.

[113] M. Greicius. "Resting-state functional connectivity in neuropsychiatric disorders." *Current Opinion in Neurology*, 21(4):424–430, 2008.

[114] R. Kendell and A. Jablensky. "Distinguishing between the validity and utility of psychiatric diagnoses." *American Journal of Psychiatry*, 160(1):4–12, 2003.

[115] S. R. Kay, A. Flszbein, and L. A. Opfer. "The positive and negative syndrome scale (PANSS) for schizophrenia." *Schizophrenia Bulletin*, 13(2):261, 1987.

[116] M. Dold, M. Tardy, M. T. Samara, C. Li, S. Kasper, and S. Leucht. "Are all first-generation antipsychotics equally effective in treating schizophrenia? A meta-analysis of randomised, haloperidol-controlled trials." *World Journal of Biological Psychiatry*, 17(3):210–220, 2016.

[117] S. Leucht, K. Komossa, C. Rummel-Kluge, C. Corves, H. Hunger, F. Schmid, C. A. Lobos, S. Schwarz, and J. M. Davis. "A meta-analysis of head-to-head comparisons of second-generation antipsychotics in the treatment of schizophrenia." *American Journal of Psychiatry*, 166(2):152–163, 2009.

[118] E. W. Tuplin and M. R. Holahan. "Aripiprazole, A Drug that Displays Partial Agonism and Functional Selectivity." *Current Neuropharmacology*, 15(8):1192–1207, 2017.

[119] P. B. Jones, T. R. Barnes, L. Davies, G. Dunn, H. Lloyd, K. P. Hayhurst, R. M. Murray, A. Markwick, and S. W. Lewis. "Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1)." *Archives of General Psychiatry*, 63(10):1079–1087, 2006.

[120] D. Naber and M. Lambert. "The CATIE and CUtLASS studies in schizophrenia: Results and implications for clinicians." *CNS Drugs*, 23(8):649–659, 2009.

[121] S. R. Marder, R. D. McQuade, E. Stock, S. Kaplita, R. Marcus, A. Z. Safferman, A. Saha, M. Ali, and T. Iwamoto. "Aripiprazole in the treatment of schizophrenia: Safety and tolerability in short-term, placebo-controlled trials." *Schizophrenia Research*, 61(2-3):123–136, 2003.

[122] R. Shiloh, Z. Zemishlany, D. Aizenberg, M. Radwan, B. Schwartz, P. Dorfman-Etrog, I. Modai, M. Khaikin, and A. Weizman. "Sulpiride augmentation in peo-

ple with schizophrenia partially responsive to clozapine. A double-blind, placebo-controlled study." *British Journal of Psychiatry*, 171:569–573, 1997.

[123] A. Alexander-Bloch, R. Lambiotte, B. Roberts, J. Giedd, N. Gogtay, and E. T. Bullmore. "The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia." *NeuroImage*, 59(4):3889–3900, 2012.

[124] M. A. Porter, J.-P. Onnela, and P. J. Mucha. "Communities in Networks." *Notices of the American Mathematical Society*, 56(9):1082–1097,1164–1166, 2009.

[125] S. Fortunato and D. Hric. "Community detection in networks: A user guide." *Physics Reports*, 659:1–44, 2016.

[126] D. S. Bassett, N. F. Wymbs, M. P. Rombach, M. A. Porter, P. J. Mucha, and S. T. Grafton. "Task-based core-periphery organization of human brain dynamics." *PLoS Computational Biology*, 9(9):e1003171, 2013.

[127] S. Lui, T. Li, W. Deng, L. Jiang, Q. Wu, H. Tang, Q. Yue, X. Huang, R. C. Chan, D. A. Collier, S. A. Meda, G. Pearlson, A. Mechelli, J. A. Sweeney, and Q. Gong. "Short-term effects of antipsychotic treatment on cerebral function in drug-naive first-episode schizophrenia revealed by "resting state" functional magnetic resonance imaging." *Archives of General Psychiatry*, 67(8):783–792, 2010.

[128] E. K. Towlson, P. E. Vértes, U. Müller-Sedgwick, and S. E. Ahnert. "Brain Networks Reveal the Effects of Antipsychotic Drugs on Schizophrenia Patients and Controls." *Frontiers in Psychiatry*, 10, 2019.

[129] J. P. Onnela, D. J. Fenn, S. Reid, M. A. Porter, P. J. Mucha, M. D. Fricker, and N. S. Jones. "Taxonomies of networks from community structure." *Physical Review E*, 86(3):36104, 2012.

[130] A. Zalesky, A. Fornito, I. H. Harding, L. Cocchi, M. Yücel, C. Pantelis, and E. T.

Bullmore. "Whole-brain anatomical networks: Does the choice of nodes matter?" *NeuroImage*, 50(3):970–983, 2010.

[131] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen. "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion." *NeuroImage*, 59(3):2142–2154, 2012.

[132] T. D. Satterthwaite, D. H. Wolf, J. Loughead, K. Ruparel, M. A. Elliott, H. Hakonarson, R. C. Gur, and R. E. Gur. "Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth." *NeuroImage*, 60(1):623–632, 2012.

[133] K. R. van Dijk, M. R. Sabuncu, and R. L. Buckner. "The influence of head motion on intrinsic functional connectivity MRI." *NeuroImage*, 59(1):431–438, 2012.

[134] J. Suckling, C. Long, C. Triantafyllou, M. Brammer, and E. Bullmore. "Variable precision registration via wavelets: Optimal spatial scales for inter-subject registration of functional MRI." *NeuroImage*, 31(1):197–208, 2006.

[135] A. X. Patel, P. Kundu, M. Rubinov, P. S. Jones, P. E. Vértes, K. D. Ersche, J. Suckling, and E. T. Bullmore. "A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series." *NeuroImage*, 95:287–304, 2014.

[136] A. Zalesky, A. Fornito, G. F. Egan, C. Pantelis, and E. T. Bullmore. "The relationship between regional and inter-regional functional connectivity deficits in schizophrenia." *Human Brain Mapping*, 33(11):2535–2549, 2012.

[137] D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.

[138] D. Cordes, V. M. Haughton, K. Arfanakis, G. J. Wendt, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand. "Mapping functionally related regions of brain

with functional connectivity MR imaging." *American Journal of Neuroradiology*, 21(9):1636–1644, 2000.

[139] L. Zhan, L. M. Jenkins, O. Wolfson, J. J. GadElkarim, K. Nocito, P. M. Thompson, O. Ajilore, M. K. Chung, and A. Leow. "The significance of negative correlations in brain connectivity." *Journal of Comparative Neurology*, 2017.

[140] Y. Liu, M. Liang, Y. Zhou, Y. He, Y. Hao, M. Song, C. Yu, H. Liu, Z. Liu, and T. Jiang. "Disrupted small-world networks in schizophrenia." *Brain*, 131(4):945–961, 2008.

[141] H. He, J. Sui, Q. Yu, J. A. Turner, B. C. Ho, S. R. Sponheim, D. S. Manoach, V. P. Clark, and V. D. Calhoun. "Altered small-world brain networks in schizophrenia patients during working memory performance." *PLoS ONE*, 7(6):e38195, 2012.

[142] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. D. Bullmore. "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs." *Journal of Neuroscience*, 26(1):63–72, 2006.

[143] J. Saramäki, M. Kivelä, J. P. Onnela, K. Kaski, and J. Kertész. "Generalizations of the clustering coefficient to weighted complex networks." *Physical Review E*, 75(2):27105, 2007.

[144] M. E. J. Newman and M. Girvan. "Finding and evaluating community structure in networks." *Physical Review E*, 69:26113, 2004.

[145] M. E. J. Newman. "Finding community structure in networks using the eigenvectors of matrices." *Physical Review E*, 74:36104, 2006.

[146] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. "Fast unfolding of communities in large network." *Journal of Statistical Mechanics: Theory and Experiment*, (P10008), 2008.

[147] M. Bazzi, M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison. "Community Detection in Temporal Multilayer Networks, with an Application to

Correlation Networks." *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 14(1):1–41, 2016.

[148] J. Reichardt and S. Bornholdt. "Statistical mechanics of community detection." *Physical Review E*, 74(1):1, 2006.

[149] S. H. Lee, M. D. Fricker, and M. A. Porter. "Mesoscale analyses of fungal networks as an approach for quantifying phenotypic traits." *Journal of Complex Networks*, 5(1):145–159, 2017.

[150] H. Schütze, C. D. Manning, and P. Raghavan. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[151] K. Mueller, J. Lepsien, H. E. Möller, and G. Lohmann. "Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates." *Frontiers in Human Neuroscience*, 11(28):7900–7905, 2017.

[152] N. J. Kopell, H. J. Gritton, M. A. Whittington, and M. A. Kramer. "Beyond the connectome: The dynome." *Neuron*, 83(6):1319–1328, 2014.

[153] J. M. Buldú and M. A. Porter. "Frequency-based brain networks: From a multiplex framework to a full multilayer description." *Network Neuroscience*, 2(4):418–441, 2017.